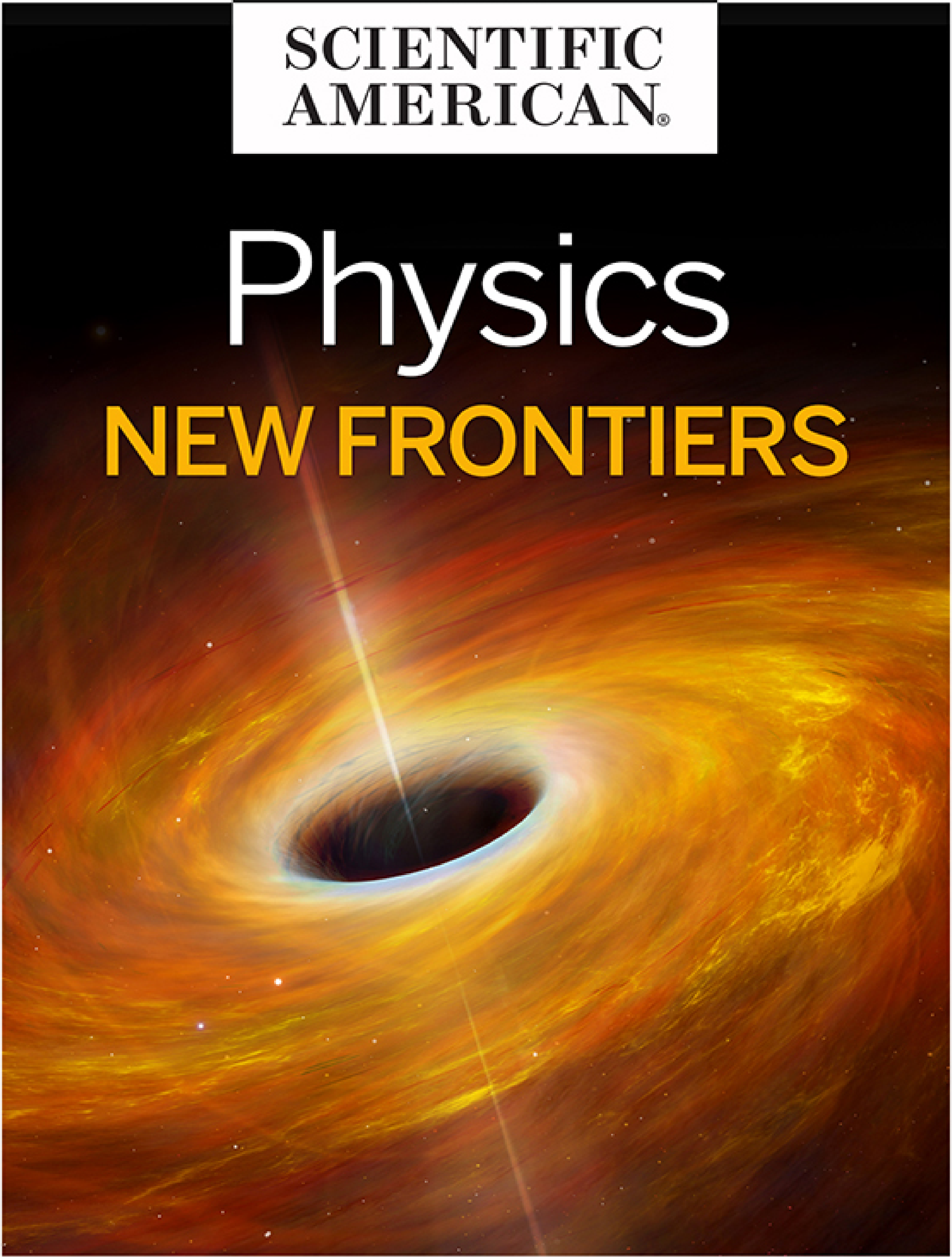
EBOOKS

SCIENTIFIC AMERICAN®

# Physics

## NEW FRONTIERS



# Physics: New Frontiers

From the Editors of Scientific American

Cover Image: Mark Garlick/Science Photo Library/Getty Images

**Letters to the Editor**

Scientific American
One New York Plaza
Suite 4500
New York, NY 10004-1562
or editors@sciam.com



Published by Scientific American
www.scientificamerican.com

ISBN: 978-1-2501-2149-3

SCIENTIFIC AMERICAN®

# PHYSICS: NEW FRONTIERS

From the Editors of Scientific American

# Table of Contents

4.2
Limits of Intelligence
*by Douglas Fox*

# Better Than Sci-Fi

Science fiction has imagined some pretty wild ideas about the universe and our place in it. Parallel or alternative universes have been a recurrent theme in *Star Trek*. *In Interstellar*, an astronaut explores hidden extra dimensions. *The Matrix* and other movies have depicted ordinary life on Earth as nothing but a holographic or mental projection. But these imaginings all seem downright tame compared with the mind-bending science now coming out of physics and astronomy.

The weirdness begins at the moment of creation. We have long thought of the big bang as the beginning of time, but theorists now have an idea of what might have come before. Could our 3-D universe have sprung from the formation of a black hole in a 4-D cosmos? The math says: maybe.

Then again, black holes might be quite different from the conventional picture of a pointlike singularity surrounded by an invisible event horizon—the boundary beyond which escape is impossible. As physicist Joseph Polchinski explains, the laws of quantum physics suggest instead that black holes may literally be large, spherical holes devoid of space and time. If so, then event horizons might actually be solid shells. And they should be quite visible, seething with an intense, instantly lethal fire of high-energy particles.

Increasingly, it seems as though hardly anything out there is what it first appears to be. The first stars to form may have been so huge they arguably deserve a new label. Our own Milky Way galaxy has a gigantic dumbbell of glowing gas skewering its center that no one noticed until recently. Even our most fundamental notions of what reality is are now up for debate—although there may be less dramatic ways to interpret the bizarre behaviors seen in the quantum realm.

Much of this new science is made possible by technological innovations. They include 5,000-odd sensors frozen deep within a cubic kilometer of

crystal-clear ice in Antarctica—an array that might shine light on the puzzling nature of dark matter. Scientists are also dissecting molecules with the most powerful x-ray laser in the world and using a 570-megapixel camera to scan the heavens for clues to the mystery of dark energy. And don’t forget the most useful tool of all for physics: mathematics.

Advanced physics can seem abstract, but it does connect to everyday life. Living things, too, must abide by the rules of quantum mechanics. And physics may eventually set an upper limit to human intelligence. In the meantime, it makes us smarter.

--W. Wayt Gibbs
Book Editor

# SECTION 1

# Amazing Astrophysics

# The Black Hole at the Beginning of Time

by Niayesh Afshordi, Robert B. Mann, and Razieh Pourhasan

In his allegory of the cave, Greek philosopher Plato described prisoners who have spent their entire lives chained to the wall of a dark cavern. Behind the prisoners lies a flame, and between the flame and prisoners parade objects that cast shadows onto a wall in the prisoners' field of view. These two-dimensional shadows are the only things that the prisoners have ever seen—their only reality. Their shackles have prevented them from perceiving the true world, a realm with one additional dimension to the world that they know, a dimension rich with complexity and—unbeknownst to the prisoners—capable of explaining all that they see.

Plato was on to something.

We may all be living in a giant cosmic cave, created in the very first moments of existence. In the standard telling, the universe came into being during a big bang that started from an infinitely dense point. But according to recent calculations that we have carried out, we may be able to track the start of the universe back to an era before the big bang, an era with an additional dimension of space. This protouniverse may have left visible traces that upcoming astronomical observations could uncover.

The universe appears to us to exist in three dimensions of space and one of time—a geometry that we will refer to as the "three-dimensional universe." In our scenario, this three-dimensional universe is merely the shadow of a world that has four spatial dimensions. Specifically, our entire universe came into being during a stellar implosion in this suprauniverse, an implosion that created a three-dimensional shell around a four-dimensional black hole. Our universe is that shell.

Why would we postulate something that sounds, on the face of it, so absurd? We have two reasons. First, our ideas are not idle speculation—they are firmly grounded in the mathematics that describe space and time.

Over the past couple of decades physicists have developed a rich theory of holography, a set of mathematical tools that al lows them to translate descriptions of events in one dimension to the physics of a different dimension. For example, researchers can solve relatively straightforward equations of fluid dynamics in two dimensions and use those solutions to understand what is go ing on in a much more complicated system—for example, the dynamics of a three-dimensional black hole. Mathematically, the two descriptions are interchangeable—the fluid serves as a perfect analogue for the extraordinary black hole.

The success of holography has convinced many scientists that more is at work here than a simple mathematical transformation. Perhaps the boundaries between dimensions are less stable than we thought. Perhaps the rules of the cosmos are written in another set of dimensions and translated into the three we perceive. Perhaps, like Plato's prisoners, our personal circumstances have tricked us into believing the world is three-dimensional when in fact a deeper understanding of what we perceive will come only when we look for explanations in the fourth dimension.

There is a second reason that our four-dimensional universe is worth thinking about. A close study of this universe could help us understand deep questions about the origin and nature of the cosmos. Consider, for example, the big bang, the primordial flash that brought our universe into existence. Modern cosmology holds that the big bang was immediately followed by "inflation"—a period of rapid expansion of space in which the early universe in creased its volume by a factor of $10^{78}$ (or more). Yet this expansion provides no insight into what caused the big bang. Our four-dimensional universe, in contrast, gives us an answer to the ultimate mystery: Where did the universe come from?

## The Known and Unknown Cosmos

Our investigations into the four-dimensional universe came about because of the problems that we have had contemplating the three-dimensional one. Modern cosmology has been fantastically successful, but its successes belie

deep and complex mysteries that may lend themselves to a holographic explanation.

Cosmologists can describe the history of the entire universe—from the present day all the way back to a fraction of a fraction of a second after the big bang—using only a few equations (chief among them the ones provided by Albert Einstein) and five independent numbers, or parameters. These parameters include the densities of ordinary matter, dark matter and dark energy (more on these in a moment), along with the amplitude and shape of quantum fluctuations in the early universe. This model—the Lambda Cold Dark Matter (Λ-CDM) cosmological paradigm—describes hundreds (if not thousands) of observational data points, covering scales from a million light-years to 10 billion light-years across, right up to the edge of our observable universe.

But these observational successes do not mean our task is complete. The story of the universe is pocked with troublesome holes. We are confronted by fundamental questions about the nature of the cosmos—problems that we have not, as of yet, been able to answer.

***Problem 1: We don't understand the five parameters.***

We do not have a satisfactory explanation for the origin of the five parameters of the Λ-CDM model, some of which must be very precisely chosen to agree with observations. Consider the density of matter and energy in the universe. Only a few decades ago astronomers believed that ordinary matter—the elements that make up the periodic table—would be the dominant form of mass-energy. Cosmological observations have radically revised this picture (and secured three Nobel Prizes along the way). We now know that the density of ordinary matter is only 5 percent of the universe's total energy density.

Another 25 percent comes in the form of dark matter, an unknown form of matter whose existence is inferred from its gravitational attraction. And 70 percent of the universe is made of dark energy, the mysterious stuff that is causing the expansion rate of our universe to speed up instead of slowing because of gravitational attraction. What are dark matter and dark energy, and why do they make up 25 and 70 percent of the universe, respectively? We do not know.

Perhaps answers would come if we better understood the big bang—the abrupt origin of space and time in a hot plasma of radiation and particles at a temperature above $10^{27}$ degrees. It is very difficult to imagine how a situation like the universe in the moments after the big bang could lead to what we observe today—a cosmos of nearly uniform temperature and with a flat, large-scale spatial geometry (in which the angles of triangles sum up to 180 degrees).

Cosmic inflation might be the best idea we have for understanding the large-scale structure of the universe. Inflation would tend to “flatten” the universe, smoothing out any curved regions of spacetime, and bring it to a uniform temperature. Like a cosmic magnifier, inflation also amplifies tiny quantum fluctuations in energy density to cosmic size during this process. These fluctuations in turn become the seeds for the growth of structures such as galaxies, stars, planets and even living organisms such as ourselves.

Inflation is generally regarded as a very successful paradigm. For decades cosmologists have been checking on inflation’s predictions by observing the cosmic microwave background (CMB) radiation, a cosmic record of density fluctuations in the early universe. Recent observations by the Planck satellite confirm that our universe is flat (or very nearly so) and that it is uniform to better than one part in 60,000—as predicted by inflation. Furthermore, the observed amplitude and shape of primordial matter fluctuations are in broad agreement with how we would expect inflation to magnify the quantum vacuum.

---

## Before the Big Bang

In the standard story, the big bang began with a singularity, an infinitely dense point that gave rise to the entire universe. Singularities are unpredictable, however; the laws of physics break down there, and there is no reason to think that one would create the world we see. Instead the authors postulate that the universe began when a star in a four-dimensional universe collapsed to form a black hole. Our universe would be protected from the singularity at the heart of this black hole by a three-dimensional event horizon. Here we depict the process in 3-D because it is difficult to illustrate what a 4-D cosmos would look like.

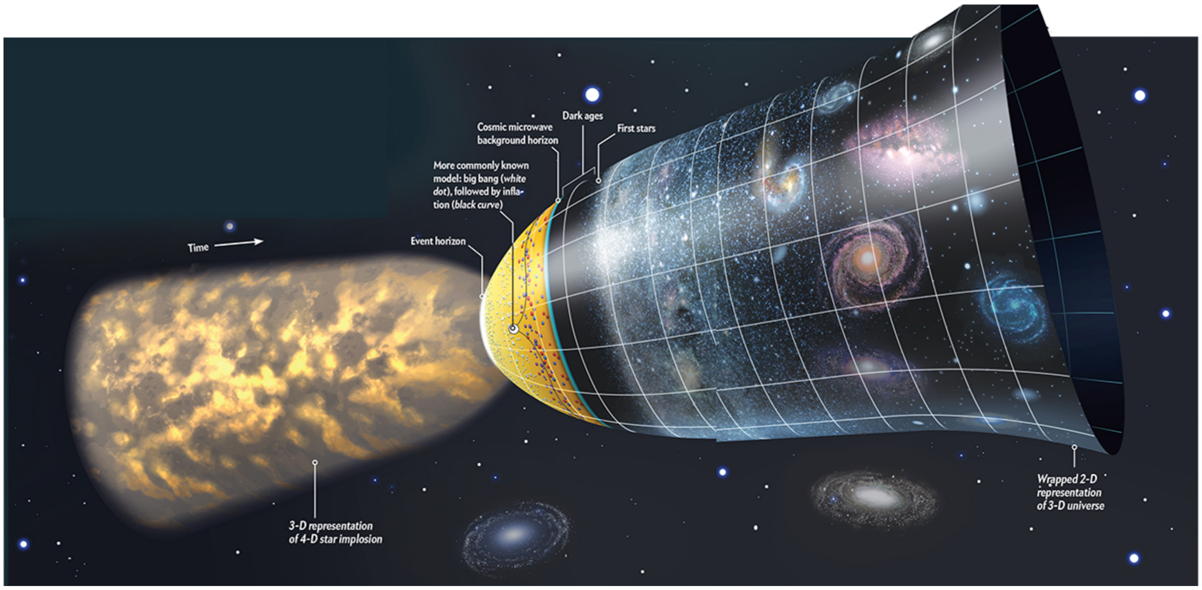



*Illustration by George Retseck*

***Problem 2: We don't understand inflation completely.***

We might ask what drove this inflation, which took a lot of energy. We imagine that, shortly after the big bang, the universe was filled with energy that takes the form of a hypothetical particle called the inflaton (pronounced "IN-flah-tahn"). The Higgs particle, recently discovered by the Large Hadron Collider at CERN near Geneva, shares many properties with, and is a possible candidate for, the proposed inflaton. The inflaton could be responsible both for early accelerated expansion and for structure in our universe because the only significant density differences in the early universe are caused by the tiny quantum fluctuations in the inflaton field's energy.

But the inflaton does not solve our problems; it just pushes them back a step. The inflaton's properties, where it came from, and how to find it remain mysteries. We are not sure whether it really exists.

In addition, physicists do not understand how to naturally end inflation. If some kind of energy field drives an exponentially expanding universe, what would make that field suddenly turn off? And we lack a satisfactory description of the history of our cosmos before the inflationary era—those first trillionths of trillionths of trillionths of a second after the big bang.

***Problem 3: We don't understand how it all began.***

Cosmology's greatest challenge is understanding the big bang itself—the sudden, violent emergence of all space, time and matter from an infinitely dense point called a singularity. A singularity is an un imaginably bizarre thing, a point where space and time curve in on themselves, making it impossible to distinguish the future from the past. All the laws of physics break down. A singularity is a universe without order or rules. Out of a singularity could come anything that might logically exist. We have no reason to think that a singularity would generate a universe as ordered as the one we see.

We would expect the emergence of a universe from a singularity to be unthinkably chaotic, marked by huge temperature fluctuations from point to point. Furthermore, the magnifying power of inflation might be expected not to smooth everything out. In fact, if these fluctuations are too large, inflation may never get a chance to begin. The problems of a singularity cannot be solved by inflation alone.

Singularities are strange, but not unfamiliar. They also form at the centers of black holes, those collapsed remains of giant stars. All stars are nuclear furnaces that fuse lighter elements (primarily hydrogen) into heavier ones. This process of nuclear fusion powers a star for most of its life, but eventually the star exhausts all its nuclear fuel, and gravity takes over. A star at least 10 times more massive than our sun will collapse on itself before exploding as a supernova. If the star is even larger—15 to 20 solar masses or more—the supernova will leave behind a dense core that goes into a runaway collapse, contracting into a point of zero size—a black hole.

Black holes can be thought of as regions of space from which not even light can escape. Because the speed of light is the maximum speed attainable by any form of matter, the boundary of a black hole—a two-dimensional surface called the event horizon—is a point of no return: once stellar matter (or anything else) falls within this boundary, it is cut off from the rest of the universe and inexorably pulled toward the singularity at the center.

As with the big bang, the laws of physics break down at this singularity as well. *Unlike* the big bang, however, a black hole is surrounded by an event horizon. This surface acts like armored wrapping paper—it prevents any information about the singularity from leaking out. The event horizon of the

black hole shields outside observers from the singularity's catastrophically unpredictable effects.

The event horizon effectively renders the singularity impotent, making it possible for the laws of physics to describe and predict all that we observe. Seen from a distance, a black hole would appear to be a simple, smooth and uniform structure, described only by its mass and angular momentum (and electric charge if it has any). Al though physicists have recently raised some interesting questions about whether this conventional picture is consistent with quantum physics, the working assumption in cosmology is that black holes are cloaked by their event horizons.

In contrast, the big bang singularity (as commonly understood) is not cloaked. It has no event horizon. We would like to have a way to shield ourselves from the big bang's singularity and its catastrophic unpredictability, perhaps by something akin to an event horizon.

We have proposed one such scenario. It turns the big bang into a cosmic mirage. Our picture cloaks the singularity at the big bang just as an event horizon cloaks the singularity at the heart of a black hole. The cloaking protects us from the singularity's mercurial and nefarious effects.

**Extradimensional Collapse**

Such a cloak would differ from an ordinary event horizon in one critical way. Because we perceive that our universe has three spatial dimensions, the event horizon that cloaks the singularity at the heart of the big bang must also have three spatial dimensions—not just two. If we imagine that this event horizon also came about as a result of a cosmic collapse—just as a black hole's two-dimensional event horizon is formed by the collapse of a three-dimensional star—then the collapse would have to have taken place in a universe that had four spatial dimensions.

This kind of extradimensional scenario, in which the number of dimensions of space exceeds the obvious three, is an idea almost as old as general relativity itself. It was originally proposed by Theodor Kaluza in 1919 and expanded by Oskar Klein in the 1920s. Their idea was largely forgotten for more than half a century before being picked up by physicists studying string theory in the 1980s. More recently, scientists have used it to build a cosmology of so-called brane worlds.

The basic idea of a brane world is that our three-dimensional universe is a subuniverse embedded in a larger space of four or more spatial dimensions. The three-dimensional universe is called a brane, and the larger universe is called the bulk. All known forms of matter and energy are stuck to our three-dimensional brane like a movie projected on a screen (or the shadow reality for Plato's prisoners in the cave). The exception is gravity, which permeates all of the higher-dimensional bulk.

Let's think about the bulk suprauniverse of four spatial dimensions that may have existed before the big bang. We can imagine that this bulk universe was filled with objects such as four-dimensional stars and four-dimensional galaxies. These higher-dimensional stars might run out of fuel, just as our three-dimensional stars do, and collapse into black holes.

What would a four-dimensional black hole look like? It would also have an event horizon, a surface of no return from which no light could escape. But instead of a two-dimensional surface, as we have in ordinary black holes, a four-dimensional black hole would generate an event horizon with three spatial dimensions.

Indeed, by modeling the collapsing death of a four-dimensional star, we find, under a certain set of assumptions, that material ejected from the stellar collapse can form a slowly expanding three-brane surrounding this three-dimensional event horizon. Our universe is this three-brane—a hologram of sorts for a four-dimensional star collapsing into a black hole. The cosmic big bang singularity becomes hidden to us, locked away forever behind a three-dimensional event horizon.

**Is This Real?**

Our model has a number of things going for it, starting with the fact that it eliminates the naked singularity that gave rise to the universe. But what of the other long-standing cosmological problems, such as the near flatness and high uniformity of the cosmos? Because the four-dimensional bulk universe could have existed for an infinitely long time in the past, any hot and cold spots in the bulk would have had plenty of time to come to equilibrium. The bulk universe would be smooth, and our three-brane universe would inherit this smoothness.

In addition, because the 4-D black hole would also ap pear to be nearly featureless, our emergent three-brane universe would likewise be smooth. The larger the mass of the four-dimensional star, the flatter the three-brane, and so the flatness of our universe is a consequence of it being residual detritus from the collapse of a heavy star.

In this way, our model of a holographic big bang resolves not only the main puzzles of uniformity and near flatness of standard cosmology without resorting to inflation but also nullifies the damaging effects of the initial singularity.

The idea may sound crazy, but there are several ways one might be able to test it. One way is by studying the cosmic microwave background radiation. Outside of our three-brane, we would expect there to be some extra four-dimensional bulk matter—something pulled close by the gravitational pull of the black hole. We can show that thermal fluctuations in this extra matter will create fluctuations on the three-brane that in turn distort the CMB by small but potentially measurable amounts. Our calculations differ from the latest data from the European Space Agency's Planck space observatory by about 3 percent. But this discrepancy may be the result of secondary effects that we are now in the process of modeling.

In addition, if the four-dimensional black hole is spinning (it is very common for black holes to spin), then our three-brane may not look the same in all directions. The large-scale structure of our universe would appear slightly different in different directions. Astronomers may also be able to find this directionality by studying subtle variations in the CMB sky.

Of course, even as the holographic big bang potentially resolves one giant question—the origin of our universe—it simultaneously raises a new set of mysteries. Foremost among them: Where did our universe's *parent* universe come from?

For an answer to this puzzle, we might again turn to Plato. When Plato's prisoners emerged from the cave, the light of the sun burned their eyes. It took them time to adjust to the brightness. At first, the prisoners were only able to make out shadows and reflections. Soon they could see the moon and the stars. Finally, they correctly concluded that the sun was "the author of all that we see"—day, night, season and shadow.

Plato's prisoners didn't understand the powers behind the sun, just as we don't understand the four-dimensional bulk universe. But at least they knew where to look for answers.

---

## Whispers from Creation or Galactic Dust?

Apart from the terrible weather hanging around from an unusually cold winter, the second week of March 2014 started like any other week. But then rumors started floating around in the cosmology community about an imminent announcement out of the Harvard-Smithsonian Center for Astrophysics. The rumors spread to Facebook, Twitter and the blogosphere by the weekend. Details began to emerge. This was not any ordinary announcement but rather the kind that, if correct, would happen once in a lifetime. It was something that most of us dreamed we could see only in a few decades if we were lucky, if at all.

The announcement from the BICEP-Keck collaboration was that an array of microwave telescopes located at the geographical South Pole had picked up startling patterns in the polarization of the cosmic microwave background. If these patterns were generated in the early universe, as the collaborators suggested was likely, they would confirm a 30-year-old prediction of the cosmic inflation theory: that the simplest models of inflation can generate an observable level of gravitational waves, comparable to density or temperature fluctuations in the early universe. It would also be our first direct evidence for the quantum nature of gravity, the most outstanding puzzle in theoretical physics over the past century.

Yet in science, as in life, things are rarely as simple as they first appear. For example, the simple inflationary models that predict observable levels of gravitational waves also suggest that hints of these waves should have been seen in the temperature fluctuations observed by the European Space Agency's Planck satellite. But they were not! Furthermore, microwave emission from dust in our galaxy tends to be polarized, which could confuse BICEP-Keck observations, at least to some extent.

As it turned out, further data released by the Planck team in late 2014 demonstrated that the original BICEP-Keck analysis had underestimated the degree of dust contamination in their maps. Dust in the Milky Way could in fact account for all of their signal. New experiments now planned or under way will improve significantly over the next decade, however. These should shed further light on the first moments after the big bang.

— N.A., R.B.M. and R.P.

---

--Originally published: Scientific American 24(4); 4-11 (Winter 2015).

# The First Starlight

by Michael D. Lemonick

About 13.8 billion years ago, just 400,000 years or so after the big bang, the universe abruptly went dark. Before that time, the entire visible universe was a hot, seething, roiling plasma—a dense cloud of protons, neutrons and electrons. If anyone had been there to see it, the universe would have looked like a pea soup fog, but blindingly bright. Around the 400,000-year mark, however, the expanding universe cooled enough for hydrogen atoms to form at last—an event known as recombination. The fog lifted, the universe continued to cool and everything quickly faded to black. After the unimaginable brilliance of the big bang and its immediate aftermath, the cosmos entered what astronomers call the dark ages of the universe.

And dark they were. For even when the first stars started to ignite, their light shone brightest in the ultraviolet portion of the spectrum—just the kind of light that the newly formed hydrogen gas tends to absorb. The universe traded its primordial hot, bright fog for one that was cool and dark.

Eventually this fog would lift, but how it did so is a question that has long baffled astronomers. Maybe it was accomplished by the first stars, whose intense light gradually but relentlessly broke the hydrogen apart in a process called reionization. Perhaps instead the energy for reionization came from the radiation that is generated by hot gas spiraling into giant black holes.

The key to figuring out how and when reionization took place is finding the oldest objects in the universe and trying to tease out their nature and their origins. When did the first stars turn on, and what were they like? How did individual stars assemble themselves into galaxies, and how did those galaxies form the supermassive black holes that lie at the core of nearly all of them? At what point in this progression from stars to galaxies to black holes did re ionization take place? And was the process gradual or abrupt?

Astrophysicists have been asking many of these questions since the 1960s. Only recently, however, have telescopes and computer models gotten powerful enough to offer some answers. Computers have now simulated the emergence and evolution of the first stars in the universe. And telescopes are gathering telltale glimmers of light from less than half a billion years after the big bang—a time when the first galaxies were in their infancy.

**Superstars**

A decade ago astronomers believed that they had a good handle on how the first generation of stars came to be. Immediately af ter recombination, the hydrogen atoms that filled the cosmos were spread uniformly through space. In contrast, dark matter, which physicists believe to be made of invisible particles that have not yet been identified, had already begun clumping together in clouds known as halos, averaging somewhere between 100,000 and one million solar masses. Gravity from these halos sucked in the hydrogen. As the gas became increasingly concentrated and heated up, it flared into light, creating the first stars in the universe.

In principle, this initial generation of giant stars, known to astronomers as Population III stars, could have broken up the gaseous veil of hydrogen atoms and reionized the universe. But much depends on the exact characteristics of these stars. If they were not bright enough or did not live long enough, they would not be capable of finishing the job.

The characteristics of these stars depend strongly on their size. As of a decade ago, astronomers believed that they would be uniformly gigantic, each with roughly 100 times the mass of the sun. The reason: As a clump of gas tries to collapse under gravity, it heats up. The heat creates so-called radiation pressure that opposes gravity; unless the star can shed some of this heat, the collapse will stall.

The first stars were made mostly of hydrogen, which is relatively terrible at shedding heat. (Stars like our sun also have small but critical traces of elements such as oxygen and carbon, which help them to cool.) As a result, a protostar in the early universe would continue to accumulate hydrogen gas, but the high pressure would prevent it from forming a dense core that would burst into a fusion reaction—one that drives much of the surrounding

gas back out into space. The star would just gorge itself on more and more gas until it built a massive, diffuse core.

Now, however, says Thomas Greif, who created some of the most sophisticated simulations of early star formation while working as a postdoctoral fellow at Harvard University, “things look a bit more complicated.” These newest simulations include not just gravity but also equations describing the feedbacks generated by increasingly pressurized hydrogen as the gas collapses. It turns out that the collapse of a hydrogen cloud can play out in many different ways. In some cases, the first stars could have been up to a million times as massive as the sun. In others, the collapsing cloud would have fragmented, creating several stars of just a few tens of solar masses.

These enormous size differences imply correspondingly huge variations in the possible lifetimes of the first stars—and therefore in the onset and duration of the epoch of reionization. Giant stars of 100 solar masses or more are the rock ’n’ rollers of astronomy: they live fast and die young. Smaller stars would churn through their nuclear fuel more slowly, implying that if stars were responsible for reionization, the process would have played out over many hundreds of millions of years.

---

## The First Billion Years

Just 380,000 years after the big bang—a blink on cosmic scales—the universe cooled to the point where hydrogen atoms could form, and all went dark. By about one billion years later, the universe was fully “reionized”—radiation had blown the atoms apart, clearing the way for light to shine again. But which objects powered the reionization—were they stars or galaxies or the black holes at the centers of quasars?

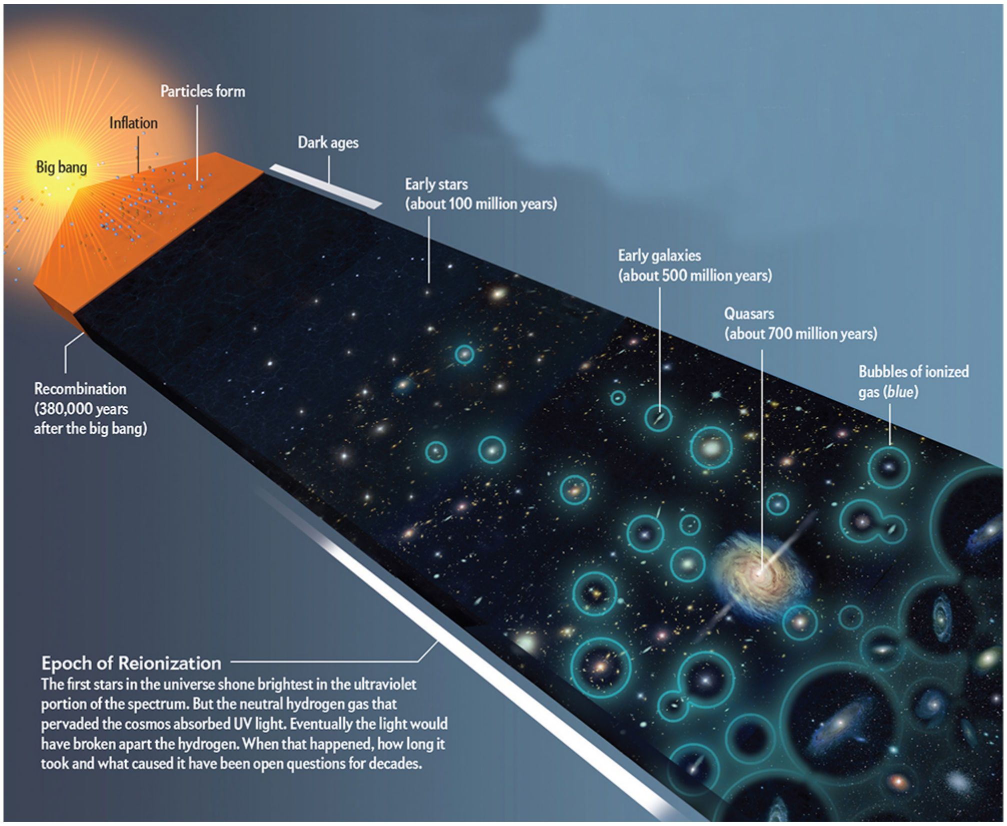



*Illustration by Moonrunner Design*

---

## Black Lights

However large they were, all of these stars would have ended their existence in fiery supernovae before collapsing into black holes. And these black holes—perhaps more than the stars from which they came—may have then fueled the engine of reionization.

Black holes swallow nearby gas voraciously, and as the gas falls in, gravity compresses and heats it to temperatures of millions of degrees. The gas is so hot that while most of it eventually disappears into the black hole, some spews back out into space in the form of jets, which shine so brightly that the light can be seen halfway across the cosmos. We call these beacons quasars.

From the 1960s through the 1990s, quasars were really the only way to probe the early universe. At first, astronomers had no clue what they were. Quasars looked like nearby stars but had huge redshifts—a reddening of their light caused by the expansion of the universe. The impressive redshifts indicated that quasars were vastly farther away than any standalone star

could possibly be and were thus vastly brighter as well. The first one ever found, 3C 273, had a redshift of 0.16, indicating that its light had begun traveling about two billion years ago.

“Then, very quickly,” says Princeton University astrophysicist Michael A. Strauss, “people found quasars up to redshift 2”—a look-back time of more than 10 billion years. By 1991 Maarten Schmidt, James E. Gunn and Donald P. Schneider, working together at Palomar Observatory in California, had found a quasar with a redshift of 4.9, dating to 12.5 billion years before the present, or just a billion years and change after the big bang.

Yet analyses of the redshift 4.9 quasar found no evidence that its light was being absorbed by neutral hydrogen. Apparently the universe had already been reionized by the time the light from this quasar had begun its journey to Earth.

For most of the 1990s, no one was able to find a quasar any farther away than this one. The failure was not for a lack of powerful instruments—both the Hubble Space Telescope and the Keck telescopes at Mauna Kea in Hawaii came online in the early 1990s, significantly increasing astronomers’ ability to see deep into the universe—but because quasars are rare to begin with. They erupt only from the most massive of supermassive black holes. And we can detect them only if their jets of gas happen to be aimed directly at us.

Moreover, those jets blast into existence only when a black hole is actively swallowing gas. For most black holes, that kind of activity peaked between a redshift of 2 and 3, when galaxies were more gas-rich, on average, than they are today. If you look further out than that sweet spot in cosmic time, the number of quasars drops off rapidly.

It was not until 2000, when the Sloan Digital Sky Survey began methodically searching across an enormous swath of sky with the largest digital detectors ever built until that point, that the record was shattered in earnest. “The Sloan was just fabulously successful in finding distant quasars,” says Richard Ellis , a professor of astrophysics at University College London. “They found something like 40 or 50 quasars beyond a redshift of 5.5.” But the survey could not reach back much further than a

handful of quasars that it found between redshift 6 and 6.4, and even at that distance there was no sign of neutral hydrogen.

It was only with the discovery of a quasar at redshift 7.085, by the UKIRT Infrared Deep Sky Survey at Mauna Kea, that astronomers found small but significant amounts of ultraviolet-absorbing hydrogen obscuring the object's light. This quasar, known as ULAS J1120+0641, was shining about 770 million years after the big bang. It finally let astrophysicists dip a toe into the era of cosmic reionization—but just a toe because even this close to the big bang, most of the neutral hydrogen had already been destroyed.

Or maybe not. It is possible that this quasar sits in an unusually sparse region of leftover neutral hydrogen and that most other quasars at this distance would have been more completely shrouded. It is equally possible that ULAS J1120+0641 is in an especially dense region; maybe reionization was already complete pretty much everywhere else. Without more examples, astronomers cannot be sure, and the prospects of finding enough quasars at this distance to do a robust statistical study are slim.

ULAS J1120+0641 has plenty to tell astronomers anyway. For one thing, Ellis says, "the number of quasars is falling so steeply with distance that it's inconceivable that massive black holes are a major source of radiation that reionizes the universe." For another, the black hole that powers this particular quasar must have a billion suns' worth of mass to generate the amount of energy that makes it visible from so far away. "It's almost impossible to understand how it could have formed in the limited time that the universe had up to that point," Ellis says.

Nevertheless, form it did. Abraham Loeb, chair of Harvard's astronomy department, points out that if a first-generation star of 100 solar masses collapsed into a black hole a couple of hundred million years after the big bang, it could conceivably have grown to quasar proportions in the available time if conditions were just right. "But you would need to feed the black hole continuously," he says, and it is hard to imagine how you could do that. "They shine so brightly, they produce so much energy, that they expel the gas out of their vicinity." Without a nearby supply of gas, the quasar goes temporarily dark, allowing gas to accumulate again until it can flare back into life—and blow away its fuel supply once more. As this

cyclical process repeats, Loeb says, “the black hole is able to grow only for a fraction of the time.”

Yet black holes can also get bigger by merging with one another, and merging should accelerate their growth. In addition, Loeb and his co-authors suggested in a 2003 paper that those first black holes may have formed from stars that were not 100 solar masses, but one million solar masses. “This has become a popular idea,” Loeb says, buttressed by the recent work on star sizes and by simulations such as Greif’s. “And because these stars would shine as brightly as the entire Milky Way, you could, in principle, see them with the James Webb Space Telescope,” the massive successor to the Hubble telescope that is currently scheduled to launch in 2018.

---

Young Giants of the Universe

Why were the first stars so large? All stars execute a cosmic balancing act—gravity attempts to squeeze them as tight as possible, but the gas pressure inside the star fights against gravity and keeps the star inflated. By comparing star formation in the modern universe with star formation in the early universe, we can begin to understand why the universe’s first stars were so massive.

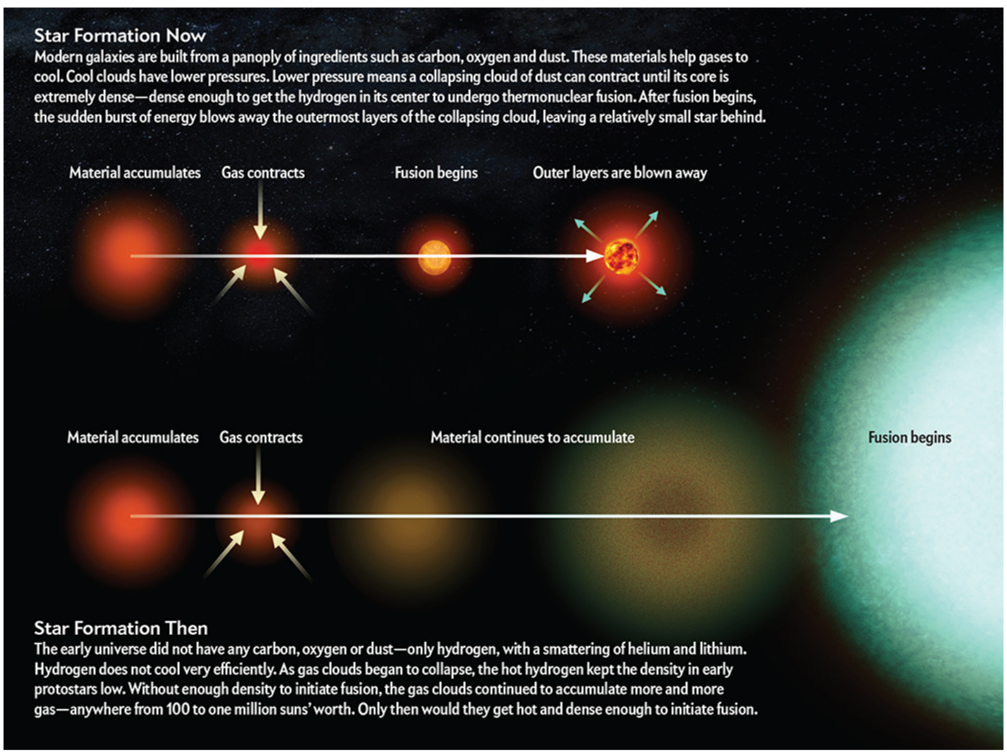



*Illustration by Moonrunner Design*

---

## Galaxy Quest

Even as the hunt for distant quasars has more or less petered out, the search for galaxies closer and closer to the big bang has taken off. Perhaps the most important triggering event was an image called the Hubble Deep Field. It was made in 1995, when Robert Williams, then director of the Space Telescope Science Institute, used a perk of the office known as "director's discretionary time" to aim Hubble at an evidently blank patch of sky and let it stare for a cumulative 30 hours or so to pick up whatever faint objects might be there. "Some very serious astronomers told him it was a waste of time," recalls former director Matt Mountain, "that he wouldn't see anything."

In fact, the telescope picked up several thousand small, faint galaxies, many of them among the most distant ever seen. Later Deep Field images—made with Hubble's upgraded, infrared-sensitive Wide Field Camera 3, which is about 35 times more effective than its predecessor was—have found even more.

"We've gone from four or five galaxies with a redshift of 7 or more to more than 100," Daniel Stark , a professor of astronomy at the University of Arizona. One of these ultra distant galaxies, described by Stark, Ellis and several co-authors in a 2012 paper, appears to be at a redshift of no less than 11.9, which dates it to fewer than 400 million years after the big bang.

Like the record-holding quasar, these young galaxies can tell astronomers plenty about the distribution of intergalactic hydrogen at the time. When observers look at their output of ultraviolet light, a significant fraction of what they would expect to see is missing, absorbed by neutral hydrogen that surrounds them. That fraction drops gradually as they look at galaxies that are further from the big bang—until, at about a billion years after the universe was born, the cosmos is fully transparent.

In short, not only did galaxies exist to provide a source for the ionizing radiation, they also reveal how the universe made the transition from neutral to fully ionized. Astronomical detectives have a smoking gun, and they have a victim. There is a catch, however. When they take the 100-odd galaxies found so far above a redshift of 7 and extrapolate across the entire sky, they do not come up with enough total ultraviolet radiation to ionize all the neutral hydrogen. The gun does not seem to have been powerful enough to do the job. The required energy could not have come from black holes,

either; there just was not enough time for that many supermassive black holes to form.

Yet the answer may be relatively straight forward. Faint as they seem to us, the galaxies we are able to see at the very edge of Hubble's vision are presumably the brightest of their epoch. There must be many more galaxies at that distance that are simply too dim to see with any existing telescope. The assumption is reasonable enough that "I think most people now believe that galaxies do most of the work in reionizing the universe," Ellis says.

### Quasar Quest

Quasars are among the brightest objects in the early universe, beacons that astronomers can spot from more than 10 billion light-years away. As light from a quasar travels through the universe toward our telescopes, two things happen: First, its light gets stretched along the way by the expansion of the universe. In addition, intervening atomic hydrogen gas absorbs some of the light. Astronomers plot the absorption of light by wave length to see how the prevalence of hydrogen gas has changed over time. They have found that isolated bubbles of ionized gas grew larger and more numerous as the universe evolved.

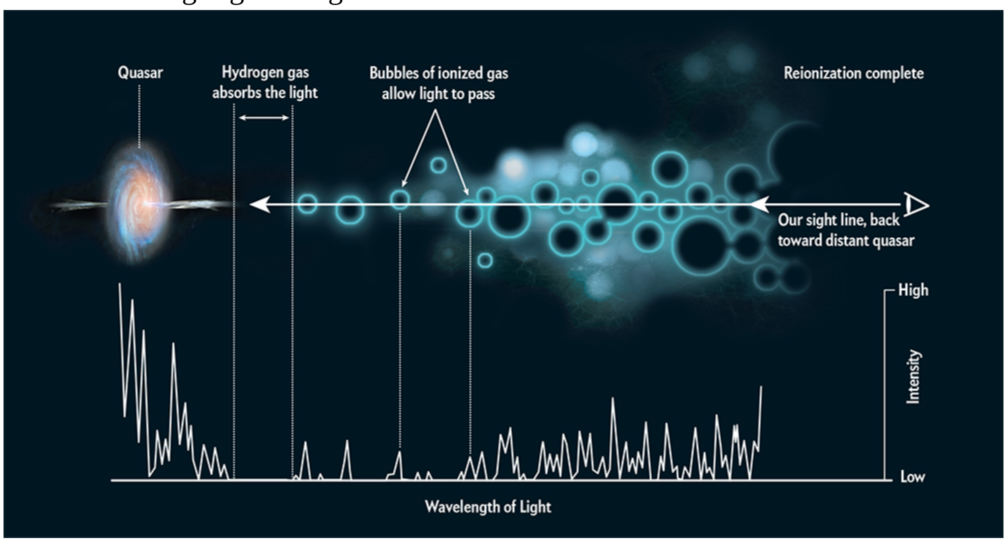



*Illustration by Moonrunner Design*

## The Einstein Card

As for what truly newborn galaxies look like and when they first turned on, "we're not there yet," Stark admits. "The galaxies we do see are fairly small, and they look much younger than galaxies that have been studied in detail [from] one billion to two billion years later." But these small, young galaxies already have as many as 100 million stars. And the mix of their colors (after you correct for the fact that their light is redshifted) suggests their stars are on average redder than one would expect to see in an infant

galaxy. “These objects,” Stark says, “look like they’ve been forming stars for at least 100 million years already. Hubble has taken us close to the precipice, to where we’ll see the first generation of stars. But it will take the James Webb Space Telescope to get us there.”

Hubble has not exhausted its options, however. The telescope itself can see only to a certain faintness limit without taking absurdly long exposures. Yet the universe has supplied its own natural lenses that can boost Hubble’s power. These so-called gravitational lenses take advantage of the fact that massive objects—in this case, clusters of galaxies—warp the space around them, distorting and sometimes magnifying the objects that lie far beyond.

In particular, says observer Marc Postman of the Space Telescope Science Institute, “we get a big amplification of any very distant galaxies that lie behind those clusters. They can be 10 or 20 times brighter than comparable unlensed galaxies.” Postman is principal investigator for the Cluster Lensing and Supernova Survey with Hubble, a program that has used the technique to identify some 250 additional galaxies between redshift 6 and 8 and a handful more that may go up to redshift 11. From what they have seen so far, the results are consistent with those coming out of the various Deep Field surveys.

Recently Hubble has been going deeper still in a project called Frontier Fields, which started in 2013. Through 2016 , Mountain and other astronomers will look for magnified images of faint, distant galaxies that lie behind six especially massive and powerful clusters. Each cluster will be observed for 140 orbits of Hubble, totaling more than 100 hours. “That will let us probe deeper into the universe than anything we’ve ever seen,” says Jennifer Lotz, the project’s lead observer.

## Burst Search

Yet another kind of cosmic beacon, meanwhile, could ultimately prove to be an even better probe of the early universe. When first discovered in the 1960s, gamma-ray bursts—short blasts of high-frequency radiation that pop up in random directions—were an utter mystery. Nowadays astronomers believe that many of them come from the deaths of very massive stars. As the stars collapse to form black holes, they spew jets of gamma rays out into space. When the jets slam into the surrounding clouds of gas, they trigger a

secondary, bright afterglow of visible and infrared light that can be seen by conventional telescopes.

Here is how the observations work: When the orbiting Swift Gamma-Ray Burst observatory detects a gamma-ray flash, it swivels to point its onboard telescopes at the spot. At the same time, it beams the location's coordinates to ground-based observers. If their telescopes get there before the flash fades, astronomers can measure the afterglow's redshift and thus the redshift—and age—of the galaxy where the burst went off.

What makes the technique so valuable is that gamma-ray bursts make other cosmic objects look positively feeble. "For the first few hours," says Edo Berger, a Harvard astrophysicist who specializes in the bursts, "they probably outshine galaxies by a factor of a million, and they're 10 to 100 times brighter than quasars." You do not need a long exposure with Hubble to see them. In 2009 a telescope on Mauna Kea reliably measured the redshift of one burst at 8.2, putting it at 600 million years after the big bang.

The flash was so bright, Berger says, that it could have been seen out to a redshift of 15 or even 20, which would be less than 200 million years after the big bang, close to the time the very first stars may have been shining. And it is entirely plausible, he says, that those very massive stars would be exactly the kind to produce gamma-ray bursts as they die. In fact, Berger says, there is reason to think these first-generation stars would create such energetic gamma-ray bursts that they would appear brighter than the ones discovered so far, even though they would be farther away.

Unlike quasars, moreover, which occur only in galaxies with supermassive black holes, and unlike the galaxies that Hubble can see, which are the brightest tips of a grand galactic iceberg, gamma-ray bursts are just as powerful in tiny galaxies as they are in big ones. They provide, in other words, a much more representative sample of the universe at any given time.

The downside, Berger says: 99 percent of gamma-ray bursts are pointed away from Earth, and of the remaining one per day or so that are detected by satellites, only a minuscule fraction are at a high redshift. Gathering a representative sample of extremely high redshift bursts would therefore take a decade or more, and Swift probably will not last that long, Berger says.

Ideally, he notes, someone should launch a successor satellite that could feed burst coordinates to the James Webb telescope or to the three 30-meter-class ground-based instruments that are expected to be operating within the next decade. Proposals to do so have so far failed to get the go-ahead from either NASA or the European Space Agency.

In any case, once the James Webb telescope and the next generation of gigantic ground-based telescopes begin observations, quasar hunters, galaxy surveyors and those who search for the telltale afterglows of gamma-ray bursts in other electromagnetic wavelengths will be able to catalogue much older and fainter objects than they can today. Their work will help to nail down exactly what was going on in the very early universe.

Radio astronomers, meanwhile, will be looking to a number of instruments—including the Murchison Widefield Array in Australia, the Precision Array for Probing the Epoch of Reionization in South Africa and West Virginia, the Square Kilometer Array to be built in Australia and South Africa, and the Low Frequency Array antennas located in several European countries —to map slowly disappearing clouds of neutral hydrogen during the first billion years of cosmic history.

The hydrogen itself emits radio waves. So in principle astronomers can look at those emissions at different epochs, each redshifted by a different amount, depending on how far away they are. They can then piece together snapshots of the hydrogen as it is gradually eaten away by high-energy radiation. Astronomers have used the Atacama Large Millimeter/submillimeter Array in the Chilean desert, for example, to detect carbon monoxide drifting in the space between the second-generation stars in immature galaxies a billion years after the big bang.

When cosmologists first detected the leftover electromagnetic radiation from the big bang in 1965, it galvanized them to try to understand the life history of the universe from its birth right through to the present. They are not quite there yet. But there is every reason to believe that by 2025, the 60th anniversary of that discovery, the last remaining blanks will finally be filled in.

--Originally published: Scientific American 24(4); 12-19 (Winter 2015).

# Burning Rings of Fire

by Joseph Polchinski

Falling into a black hole was never going to be fun. As soon as physicists realized that black holes exist, we knew that getting too close to one spelled certain death. But we used to think an astronaut falling past the point of no return—the so-called event horizon—would not feel anything special. Ac cording to Einstein's general theory of relativity, no signposts would mark the spot where the chance of escape dropped to zero. Anyone journeying past the horizon would just seem to fall down, down, down into a pit of blackness.

Recently, however, my colleagues and I have recast that picture in light of some new information about the effects of quantum mechanics on black holes. It now seems that our astronaut would experience something very different from Albert Einstein's prediction. Rather than falling seamlessly into the interior, the astronaut would hit a "firewall" of high-energy particles at the horizon and meet an instant demise. The wall might even mark the end of space.

Three years ago four of us, all then at the University of California, Santa Barbara—my colleague Donald Marolf, then graduate students Ahmed Almheiri and James Sully, and I (now known by the acronym AMPS)—arrived at this conclusion after using ideas from string theory to take a closer look at the physics of black holes, particularly at an interesting argument put forward in the 1970s by Stephen Hawking of the University of Cambridge. Hawking had identified a deep conflict between the predictions of quantum theory and relativity in these extreme environments. According to his reasoning, either quantum mechanics or Einstein's depiction of space time is flawed. The battle over which view is correct has swung back and forth ever since.

As with Hawking’s original claim, our recent firewall proposal has raised a storm of disbelief, and no satisfactory alternative has yet emerged. If quantum mechanics is to be trusted, firewalls are the consequence. Yet their existence raises theoretical puzzles as well. It seems that physicists must give up one of our widely cherished beliefs, but we cannot agree on which one. We hope, however, that out of this confusion will come a more complete understanding of quantum mechanics and relativity—and, ideally, a way to finally resolve the apparent contradictions between these two reigning theories of physics.

### The Singularity

General relativity, which gave birth to the very concept of black holes, derives its picture of these mysterious entities and their event horizons from an understanding of gravity’s effect on space and time. According to the theory, if enough mass comes together, gravity’s pull will cause it to start collapsing. Nothing can stop this process until all the mass is compressed into a single point where spacetime is infinitely dense and infinitely curved, called the singularity—in other words, a black hole.

Any space travelers who pass the black hole’s event horizon boundary will be unable to escape the gravitational pull and will soon be drawn into the singularity. Even light, once it is past the horizon, cannot escape. The singularity is a very dramatic place, but the horizon itself is supposed to be unremarkable, according to what is called the equivalence principle of general relativity; individuals falling freely into a black hole will see the same physical laws as anywhere else as they cross the horizon. Theorists are fond of saying that the entire solar system could be falling into a giant black hole right now, and we would not experience anything out of the ordinary.

### Hawking's Puzzles

The challenges Hawking posed to the traditional picture of black holes began in 1974, when he considered a strange prediction of quantum mechanics. According to this theory, pairs of particles and their antimatter counterparts constantly pop into existence and then disappear almost at once. If such fluctuations happen just outside the horizon of the black hole, Hawking showed, the pair could separate. One would fall into the

singularity, and the other would escape from the black hole and carry away some of its mass. Eventually the black hole's entire mass could be depleted through this process, termed Hawking evaporation.

For black holes found in nature, evaporation is unimportant: these black holes add mass at a much more rapid rate from gas and dust falling in than they lose to radiation. But for theoretical purposes, we can investigate what would happen if a black hole were completely isolated and we had enough time to watch the full process of evaporation. By pursuing such a thought experiment, Hawking revealed two apparent contradictions between general relativity and quantum mechanics.

**The entropy problem.** In pondering the isolated black hole, Hawking noted that the light spectrum of the eponymous radiation streaming away from it would look the same as that of a radiating hot body, meaning that the black hole has a temperature. In general, temperature arises from the motion of atoms inside objects. The thermal nature of Hawking radiation, then, suggested that the black hole should have a microscopic structure made of some kind of discrete building blocks, or bits.

The late physicist Jacob D. Bekenstein of the Hebrew University of Jerusalem had also reached this conclusion two years earlier by en gaging in thought experiments involving throwing things into black holes. The work of Bekenstein and Hawking gives a formula for the number of bits, a measure known as the black hole entropy. Entropy is a gauge of disorder, which becomes greater as the number of states that an object can have grows. The larger the number of bits in a black hole, the more possible arrangements they can have and the greater the entropy.

In contrast, general relativity describes a black hole as having a smooth geometry and indicates that every black hole of given mass, spin and charge should be exactly the same: in the words of the late physicist John Wheeler of Princeton University, "Black holes have no hair." So here is a contradiction: relativity says no hair, whereas quantum mechanics says black holes have a large amount of entropy, meaning some microscopic structure, or hair.

**The information paradox.** Hawking evaporation also gives rise to a challenge to quantum theory. According to Hawking's calculation, the

particles that escape from a black hole do not depend at all on the properties of the material that went into the hole—usually a massive star that collapsed. For example, we could send a note with a message into the black hole, and there would then be no way to reconstruct the message from the final particles that would emerge. Once the note passed through the horizon, it could not influence anything that came out later, because no information can escape from the interior. In quantum mechanics, every system is described by a formula called the wave function, which encodes the chances that the system will be in any particular state.

In Hawking's thought experiment, the loss of information means that we have no way to predict the wave function of Hawking radiation based on the properties of the mass that went into the black hole. Information loss is forbidden by quantum mechanics, so Hawking concluded that the laws of quantum physics had to be modified to allow for such loss in black holes.

You might be saying to yourself, "Of course, black holes destroy information—they destroy everything that enters them." But compare what happens if we simply burned the note. The message would certainly be scrambled, and it would be impractical to reconstruct it from the smoke. But the process of burning is described by ordinary quantum mechanics, applied to the atoms in the note, and the quantum description of the smoke would be a definite wave function that would depend on the original message. In theory, then, the message could be reconstructed through the wave function. In the case of black holes, however, there would be no definite wave function for the resulting radiation.

Based on this analogy, many theorists concluded that Hawking was wrong, that he had mistaken the scrambling of information for actual information loss. Further, some argued, if information can be lost, then it will not just happen in the exotic situation of black hole evaporation but everywhere and all the time—in quantum physics, anything that can happen will happen. If Hawking were right, we would see the signs in everyday physics, probably including severe violations of the law of conservation of energy.

Hawking's argument, though, stands up to simple objections. Unlike burning paper, black holes have horizons beyond which information cannot escape. Thus, we seem to have a sharp paradox: either modify quantum

mechanics to allow information loss or modify relativity to allow information to escape from the black hole interior.

There is a third possibility: perhaps the black hole does not evaporate completely but instead ends up as a microscopic remnant containing all the information of the star that created it. This “solution” has its own difficulties, however. For example, such a small object containing so much information would violate the Bekenstein-Hawking idea of entropy.

**Branes and Holograms**

String theory is one attempt to rectify some of the problems that arise when relativity and quantum mechanics collide, as in the case of black holes. This theory replaces the pointlike particles of previous theories with tiny loops or strands of string; these strings manage to eliminate some of the mathematical difficulties that arise when quantum mechanics and relativity are combined. Replacing points with strings does not, however, immediately change the black hole story.

A break came in 1995, when I was looking at another kind of thought experiment, studying strings in small spaces. Building on work that I and several others had done a few years earlier, I showed that string theory, as it was then understood, was not complete. Rather it required the existence of objects with more dimensions than the three of space and one of time we are familiar with. In black holes these higher-dimensional objects, called D-branes, would be tiny—wrapped up in hidden dimensions too small for us to detect. The next year Andrew Strominger and Cumrun Vafa, both now at Harvard University, showed that strings and D-branes together provide the precise number of bits to account for black hole entropy, at least for certain very symmetrical black holes. The entropy puzzle was partly solved.

Then, in 1997, Juan Maldacena, now at the Institute for Advanced Study in Princeton, N.J., came up with a way around the next problem: the information loss. His solution has been called the Maldacena duality. A duality is a surprising equivalence between two things that seem very different. Maldacena’s duality shows that the mathematics of a theory combining quantum mechanics and gravity—a quantum theory of gravity—based on string theory is equivalent to the mathematics of an ordinary quantum theory under a special set of circumstances.

In particular, the quantum physics of a black hole is equivalent to that of an ordinary gas of hot nuclear particles. It means that spacetime is profoundly different from what we perceive, more like a three-dimensional hologram projected from a more fundamental two-dimensional surface of a sphere.

Using Maldacena’s duality, physicists also get a way to de scribe the quantum mechanics of black holes in the bargain. If Maldacena’s assumptions are true, then ordinary quantum laws would apply to gravity as well, and information cannot be lost. By a less direct argument, evaporating black holes cannot leave behind any remnants, so it must be that the information gets out with the Hawking radiation.

Maldacena’s duality is arguably the closest approach to unifying general relativity and quantum mechanics, and he discovered it by chasing down the black hole puzzles of entropy and information loss. It is not yet proved to be true, but it is supported by much evidence—enough that in 2004 Hawking announced that he had changed his mind about the need for black holes to lose information and publicly paid off a bet with physicist John Preskill at the International Conference on General Relativity and Gravitation in Dublin.

Physicists widely believed that no single observer would see any violation of relativity or any other laws near a black hole that lived by Maldacena’s rules, but his duality falls short in not giving a clear explanation for how information gets from the inside of a black hole to the outside. One possible workaround lies in a proposal made about 20 years ago by Leonard Susskind of Stanford University and Gerard ’t Hooft of Utrecht University in the Netherlands. Invoking a relativity principle called black hole complementarity, they argued that, in essence, anyone inside the event horizon sees the information as remaining inside, whereas external viewers see it come out. There is no contradiction because these two observers cannot communicate.

**The Firewall**

Maldacena's Duality and black hole complementarity seemed to dispel all the paradoxes, but many of the details had yet to be filled in. Three years ago my own AMPS collaboration tried to make a model of how the

combined picture would work, building on ideas of physicists Samir D. Mathur of Ohio State University and Steven Giddings of U.C. Santa Barbara (and extending, unbeknownst to us, an earlier argument of Samuel Braunstein of the University of York in England). After failing repeatedly to make a successful model, we realized that the problem ran deeper than our mathematical shortcomings. A contradiction remained.

This contradiction pops up when considering quantum entanglement—the most unintuitive quant um phenomenon and the one furthest from our experience. If particles were like dice, entangled particles would be two dice that always added to seven: if you roll the dice, and the first comes up as two, then the second will always come up as five, and so on. Similarly, when scientists measure the properties of one entangled particle, the measurement also determines the characteristics of its partner.

It is a further consequence of quantum theory that entanglement is monogamous. A particle can be fully entangled with only one system at a time. If particle B is entangled with particle A, then it cannot also be entangled with a particle C that is part of some independent system.

In the case of the black hole, think about a Hawking photon, call it "B," emitted after the black hole is at least halfway evaporated. The Hawking process implies that B is part of a pair; call its partner that falls into the black hole "A." A and B are entangled. Now suppose that information is not lost to the outside universe. Mathematically, this means B must be entangled with a combination—call it "C"—of earlier Hawking photons outside of the event horizon. But then we have polygamy!

The price of saving quantum mechanics—of having B and C entangled and nothing out of the ordinary on the outside of the black hole—is sacrificing entanglement between A and B. The Hawking photons A and B began just inside and outside the horizon when they arose as an ephemeral particle-antiparticle pair. In quantum theory, the cost of breaking this entanglement, like the cost of breaking a chemical bond, is energy. Breaking the entanglement for all the Hawking pairs implies that the horizon is a wall of high-energy particles; we called it a firewall. An infalling astronaut, rather than moving freely through the horizon, encounters something dramatic—and certainly lethal.

Finding such a large departure from general relativity—a wall of energy in a place where nothing unusual should be happening—was disturbing, but the argument was simple, and we could not find a flaw. In a sense, we had just run Hawking's original argument backward, assuming that information is not lost and seeing where that assumption would lead. We concluded that, rather than the subtle effects of complementarity, there was a drastic breakdown of general relativity. As we began to describe the argument to other physicists, most were skeptical at first but then came to share our puzzlement.

Either these strange firewalls actually exist, or it seems we must again consider letting go of some of the deeply held doctrines of quantum theory. Information may not be destroyed, but perhaps some rewriting of quantum mechanics is in store. Unfortunately, observing real black holes will not decide the issue—any radiation from a firewall would be weakened by the gravitational pull of the black hole, making the firewall very hard to see.

### The End of Space

Furthermore, if the firewall exists, what is it? One idea is that the firewall is simply the end of space. Black holes are not actually black—they glow with Hawking radiation. But they may truly be holes: bubbles inside of which spacetime cannot form because, as Marolf once remarked, "the black hole's quantum memory is full." If so, then space ends at the event horizon, which contains all of the mass of the black hole. And an infalling astronaut who hits this surface dissolves into quantum bits, which then reside on its boundary.

To avoid such bizarre scenarios, physicists have attempted to circumvent the firewall conclusion. One idea is that because Hawking radiation particle B must be entangled with both A and C, then A must be part of C: the photon behind the horizon is somehow the same bit that is encoded in the earlier Hawking radiation, even though they are in very different places. This notion is something like the original idea of black hole complementarity, but to make a concrete model of this scenario, it seems, one ends up modifying quantum mechanics again. The most radical idea, from Maldacena and Susskind, is that every pair of entangled particles is connected by a microscopic spacetime wormhole, so that large regions of

spacetime, such as the black hole interior, can be built up from large amounts of entanglement.

Hawking had proposed that general relativity works for black holes but that quantum mechanics breaks down. Maldacena concluded that quantum mechanics is unmodified but that spacetime is holographic. Perhaps the truth is somewhere in the middle.

Many other ideas have been proposed, most of which give up one long-standing principle or another, and there is no consensus as to the right direction to resolve the problems. A common question is, What do firewalls imply for real-life black holes, such as the one in the center of our Milky Way galaxy? It is too early to say.

For now investigators are excited by this newly discovered contradiction between two of the central theories of physics. Our inability to say definitively whether the firewall is real exposes a limitation in our current formulations of quantum gravity, and theoretical physicists are rethinking their basic assumptions about the workings of the universe. Out of this may come a deeper understanding of the nature of space and time and of the principles underlying all the laws of physics. Ultimately, by unraveling the quandaries at the heart of black hole firewalls, we may finally get the break we need to unify quantum mechanics and general relativity into a single working theory.

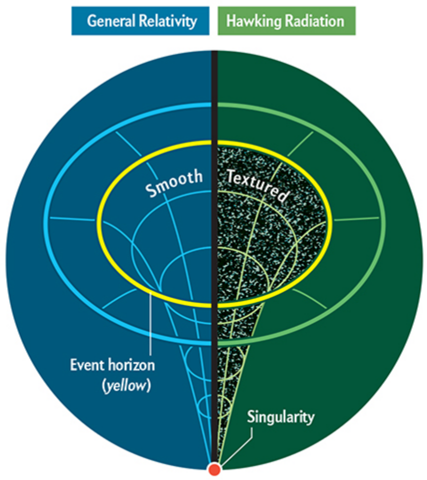
---

## Resolving Black Hole Conundrums

In 1974 Stephen Hawking showed that a small amount of radiation leaks out of black holes. According to quantum mechanics, pairs of particles and their antimatter counterparts constantly spring into existence and then disappear moments later all over the universe. Hawking noted that when a pair shows up near the horizon of a black hole, one particle could fall in while the other escapes. This phenomenon, called Hawking radiation, raises some puzzles about the laws of physics inside black holes.

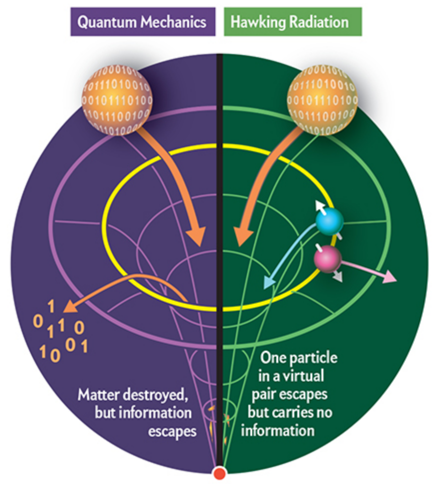
## The Entropy Problem

The radiation spectrum of Hawking emission suggests that black holes have temperatures. Heat arises from the motion of atoms and molecules within an object. The temperature of black holes implies that they have substructure, internal building blocks of some kind that can rearrange themselves. The possibility of different arrangements gives black holes a measure of disorder, or "entropy," according to the quantum-mechanical picture of Hawking radiation. Entropy is forbidden to black holes by general relativity, however, because the theory requires them to be completely smooth, lacking any substructure.




## The Information Paradox

According to the standard picture of quantum mechanics, information can never be destroyed. Even when you burn a letter, for example, the original information encoded in the atoms of the letter is preserved in the ashes and smoke. Hawking radiation, however, implies that black holes destroy the information of the matter that falls into them because the particles that escape do not depend at all on the properties of the atoms that initially fell into the hole. Stephen Hawking suggested that quantum mechanics might have to be modified to allow for information loss.




## Earlier Conjectures (*not shown*) …

In an effort to resolve these puzzles, physicists looked for new ways to combine general relativity and quantum mechanics into a coherent theory that could describe black holes. One breakthrough was string theory, which posits that particles are actually tiny loops of vibrating string. This theory appeared to solve elements of the information paradox and the entropy problem.

## … Led to Firewalls

Yet the string theory solutions eventually led to a surprising conclusion: black holes might be surrounded by firewalls—walls of high-energy particles that would violently disrupt any object that encountered them. Firewalls seem to imply a drastic breakdown of the laws of physics at the boundary of black holes and could lead to extreme conclusions, such as the possibility that firewalls mark the end of space and time altogether. If that is true, then the mass of black holes is not concentrated into an infinitely dense singularity at a pointlike center but rather spread over the two-dimensional surface of the event horizon.

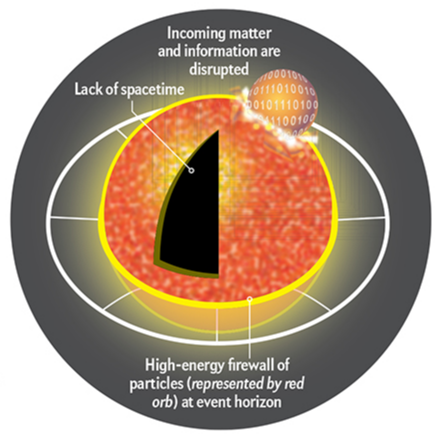



*Illustration by Jen Christiansen*

---

--Originally published: Scientific American 24(4); 20-25 (Winter 2015).

# A Brief History of Time Travel

by Tim Folger

The famous writer H.G. Wells published his first novel, *The Time Machine*, in 1895, just a few years before Queen Victoria's six-decade reign over the U.K. ended. An even more durable dynasty was also drawing to a close: the 200-year-old Newtonian era of physics. In 1905 Albert Einstein published his special theory of relativity, which upset Isaac Newton's applecart and, presumably to Wells's delight, allowed something that had been impossible under Newton's laws: time travel into the future. In Newton's universe, time was steady everywhere and everywhen. It never sped up. It never slowed down. But for Einstein, time was relative.

Time travel is not merely possible—it has already happened, though not exactly as Wells imagined. The biggest journey through time so far, according to J. Richard Gott, an astrophysicist at Princeton University, was taken by Sergei K. Krikalev. Over the course of his long career, which began in 1985, the Russian cosmonaut spent 803 days in space. As Einstein proved, time passes more slowly for objects in motion than for those at rest, so as Krikalev hurtled along at 17,885 miles an hour onboard the *Mir* space station, time did not flow at the same rate for him as it did on Earth. While in orbit, Krikalev aged $^1/_{48}$ of a second less than his fellow earthlings. Put another way, he traveled $^1/_{48}$ of a second into the future.

The time-travel effect becomes easier to see as distances stretch longer and speeds creep higher. If Krikalev left Earth in 2015 and made a round-trip to Betelgeuse—a star that is about 650 light-years from Earth—at 99.995 percent the speed of light, then by the time he returned to Earth he would be only 10 years older. Sadly, everyone he knew would be long dead because 1,000 years would have passed on Earth; it would be the year 3015.

“Time travel to the future, we know we can do,” Gott says. “It’s just a matter of money and engineering!”

Jumping a few nanoseconds—or even a few centuries—into the future is relatively straightforward, practical challenges aside. But going *backward* in time is harder. Einstein’s special theory of relativity forbade it. After another decade of work, Einstein unveiled his general theory of relativity, which finally lifted that restriction. How someone would actually travel back in time, however, is a vexing problem because the equations of general relativity have many solutions. Different solutions assign different qualities to the universe—and only *some* of the solutions create conditions that permit time travel into the past.

Whether any of those solutions describes our own universe is an open question, which raises even more profound puzzles: Just how much tweaking of fundamental physics would it take to allow backward time travel? Does the universe itself somehow prevent such journeys even if Einstein’s equations do not rule them out? Physicists continue to speculate, not because they imagine time travel to the past will ever be practical but because thinking about the possibility has led to some surprising insights about the nature of the universe we live in—including, perhaps, how it came to be in the first place.

## A New Way of Looking at Time

With his special theory of relativity, Einstein made time malleable in a way that must have pleased Wells, who presciently believed that we inhabit a universe in which three-dimensional space and time are knit together into a four-dimensional whole. Einstein arrived at his revolutionary results by exploring the implications of two fundamental ideas. First, he argued that even though all motion is relative, the laws of physics must look the same for everyone anywhere in the universe. Second, he realized that the speed of light must be similarly unchanging from all perspectives: if everyone sees the same laws of physics operating, they must also arrive at the same result when measuring the speed of light.

To make light a universal speed limit, Einstein had to jettison two commonsense notions: that all observers would agree on the measurement of a given length and that they would also agree on the duration of time’s

passage. He showed that a clock in motion, whizzing past someone at rest, would tick more slowly than a stationary clock at the person's side. And the length of a ruler moving swiftly by would shorten. Yet for anyone who was traveling at the same speed as the clock and ruler, the passage of time and the length of the ruler would appear normal.

At ordinary speeds, the time-and-space-distorting effects of special relativity are negligible. But for anything moving at a hefty fraction of the speed of light (relative to the observer), they are very real. For example, many experiments have confirmed that the decay rate of unstable particles called muons slows by an order of magnitude when they are traveling at close to the speed of light. The speeding muons, in effect, are minuscule time travelers—subatomic Krikalevs—hopping a few nanoseconds into the future.

**Gödel's Strange Universe**

Those speedy clocks and rulers and muons are all racing forward in time. Can they be thrown into reverse? The first person to use general relativity to describe a universe that permits time travel into the past was Kurt Gödel, the famed creator of the incompleteness theorems, which set limits on the scope of what mathematics can and cannot prove. He was one of the towering mathematicians of the 20th century—and one of the oddest. His many foibles included a diet of baby food and laxatives.

Gödel presented this model universe as a gift to Einstein on his 70th birthday. The universe Gödel described to his skeptical friend had two unique properties. It rotated, which provided centrifugal force that prevented gravity from crunching together all the matter in the cosmos, and thus created the stability Einstein demanded of any cosmic model. But it also allowed for time travel into the past, which made Einstein deeply uneasy. In Gödel's cosmos, space travelers could set out and eventually reach a point in their own past, as if the travelers had completed a circuit around the surface of a giant cylinder. Physicists call these trajectories in space time "closed timelike curves."

A closed timelike curve is any path through spacetime that loops back on itself. In Gödel's rotating cosmos, such a curve would circle around the entire universe, like a latitude line on Earth's surface. Physicists have

concocted a number of different types of closed timelike curves, all of which allow travel to the past, at least in theory. A journey along any of them would be disappointingly ordinary, however. Through the portholes of your spaceship, you would see stars and planets—all the usual sights of deep space. More important, time—as measured by your own clocks—would tick forward in the usual way; the hands of a clock would not start spinning backward even though you would be traveling to a location in spacetime that existed in your past.

“Einstein was already aware of the possibility of closed timelike curves back in 1914,” says Julian Barbour , an independent theoretical physicist who lives near Oxford, England. As Barbour recalls, Einstein said, “My intuition strives most vehemently against this.” The curves’ existence would create all kinds of problems with causality—how can the past be changed if it has already happened? And then there is the hoary grandfather paradox: What happens to a time traveler who kills Granddad before Granddad meets Grandma? Would the demented, now parentless traveler ever be born?

Fortunately for fans of causality, astronomers have found no evidence that the universe is rotating. Gödel himself apparently pored over catalogs of galaxies, looking for clues that his theory might be true. Gödel might not have devised a realistic model of the universe, but he did prove that closed timelike curves are completely consistent with the equations of general relativity. The laws of physics do not rule out traveling to the past.

**An Annoying Possibility**

Over the past few decades cosmologists have used Einstein’s equations to construct a variety of closed timelike curves. Gödel conjured an entire universe that allowed them, but more recent enthusiasts have warped spacetime only within parts of our universe.

In general relativity, planets, stars, galaxies and other massive bodies warp spacetime. Warped spacetime, in turn, guides the motions of those massive bodies. As the late physicist John Wheeler put it, “Spacetime tells matter how to move; matter tells spacetime how to curve.” In extreme cases, spacetime might bend enough to create a path from the present back to the past.

Physicists have proposed some exotic mechanisms to create such paths. In a 1991 paper, Gott showed how cosmic strings—infinitely long structures thinner than an atom that may have formed in the early universe—would allow closed timelike curves where two strings intersected. In 1983 Kip S. Thorne, a physicist at the California Institute of Technology, began to explore the possibility that a type of closed timelike curve called a wormhole—a kind of tunnel joining two different locations in space time—might allow for time travel into the past. “In general relativity, if you connect two different regions of space, you’re also connecting two different regions of time,” says Sean M. Carroll, a colleague of Thorne’s at Caltech.

The entrance into a wormhole would be spherical—a three-dimensional entrance into a four-dimensional tunnel in spacetime. As is the case with all closed timelike curves, a trip through a wormhole would be “like any other journey,” Carroll says. “It’s not that you disappear and are reassembled at some other moment of time. There is no respectable theory where that kind of science-fiction time travel is possible.” For all travelers, he adds, “no matter what they do, time flows forward at one second per second. It’s just that your local version of ‘forward’ might be globally out of sync with the rest of the universe.”

Although physicists can write equations that describe wormholes and other closed timelike curves, all the models have serious problems. “Just to get a wormhole in the first place, you need negative energy,” Carroll says. Negative energy is when the energy in a volume of space spontaneously fluctuates to less than zero. Without negative energy, a wormhole’s spherical entrance and four-dimensional tunnel would instantaneously implode. But a wormhole held open by negative energy “seems to be hard, probably impossible,” Carroll says. “Negative energies seem to be a bad thing in physics.”

Even if negative energy kept a wormhole open, just when you would be on the verge of turning that into a time machine, “particles would be moving through the wormhole, and every particle would loop back around an infinite number of times,” Carroll says. “That leads to an infinite amount of energy.” Because energy deforms spacetime, the entire thing would collapse into a black hole—an infinitely dense point in spacetime. “We’re not 100 percent sure that that happens,” Carroll says. “But it seems to be a

reasonable possibility that the universe is actually preventing you from making a time machine by making a black hole instead."

Unlike black holes, which are a natural consequence of general relativity, wormholes and closed timelike curves in general are completely artificial constructs—a way of testing the bounds of the theory. "Black holes are hard to avoid," Carroll says. "Closed time like curves are very hard to make."

Even if wormholes are physically implausible, it is significant that they fit in with the general theory of relativity. "It's very curious that we can come so close to ruling out the possibility of time travel, yet we just can't do it. I also think that it's annoying," Carroll says, exasperated that Einstein's beautiful theory might allow for something so seemingly implausible. But by contemplating that annoying possibility, physicists may gain a better understanding of the kind of universe we live in. And it may be that if the universe did not permit backward time travel, it never would have come into existence.

### Did the Universe Create Itself?

General relativity describes the universe on the largest scales. But quantum mechanics provides the operating manual for the atomic scale, and it offers another possible venue for closed timelike curves—one that gets at the origin of the universe.

"On a very small scale ($10^{-30}$ centimeter) you might expect the topology of spacetime to fluctuate, and random fluctuations might give you closed timelike curves if nothing fundamental prevents them," says John Friedman, a physicist at the University of Wisconsin–Milwaukee. Could those quantum fluctuations somehow be magnified and harnessed as time machines? "There's certainly no formal proof that you *can't* have macroscopic closed timelike curves," Friedman says. "But the community of people who have looked at these general questions would bet pretty heavily against it."

There is no doubt that the creation of a loop in spacetime on either a quantum scale or a cosmic one would require some very extreme physics. And the most likely place to expect extreme physics, Gott says, is at the very beginning of the universe.

In 1998 Gott and Li-Xin Li, an astrophysicist now at Peking University in China, published a paper in which they argued that closed timelike curves were not merely possible but essential to explain the origin of the universe. “We investigated the possibility of whether the universe could be its own mother—whether a time loop at the beginning of the universe would allow the universe to create itself,” Gott says.

Just as in standard big bang cosmology, Gott and Li’s universe “starts” with a bout of inflation, where an all-pervasive energy field drove the universe’s initial expansion. Many cosmologists now believe that inflation gave rise to countless other universes besides our own. “Inflation is very hard to stop once it gets started,” Gott says. “It makes an infinitely branching tree. We’re one of the branches. But you have to ask yourself, Where did the trunk come from? Li-Xin Li and I said it could be that one of the branches just loops around and grows up to be the trunk.”

A simple two-dimensional sketch of Gott and Li’s self-starting universe looks like the number “6,” with the spacetime loop at the bottom and our present-era universe as the top stem. A burst of inflation, Gott and Li theorized, allowed the universe to escape from the time loop and expand into the cosmos we inhabit today.

It is difficult to contemplate the model, but its main appeal, Gott says, is that it eliminates the need for creating a universe out of nothing. Yet Alexander Vilenkin of Tufts University, Stephen Hawking of the University of Cambridge and James Hartle of the University of California, Santa Barbara, have proposed models in which the universe does indeed arise out of nothing. According to the laws of quantum mechanics, empty space is not really empty but is filled with “virtual” particles that spontaneously pop into and out of existence. Hawking and his colleagues theorized that the universe burst into being from the same quantum-vacuum stew. But in Gott’s view, the universe is not made out of nothing; it is made out of something—itself.

## A Cosmic Chess Game

For now, there is no way to test whether any of those theories might actually explain the origin of the universe. The famed physicist Richard Feynman compared the universe to a great chess game being played by the

gods. Scientists, he said, are trying to understand the game without knowing the rules. We watch as the gods move a pawn one space forward, and we learn a rule: pawns always move one space forward. But what if we never saw the opening of a game, when a pawn can move two spaces forward? We might also assume, mistakenly, that pawns always remain pawns—that they never change their identity—until we see a pawn transformed into a queen.

"You would say that's against the rules," Gott says. "You can't change your pawn into a queen. Well, yes, you can! You just never saw a game that extreme before. Time-travel research is like that. We're testing the laws of physics by looking at extreme conditions. There's nothing logically impossible about time travel to the past; it's just not the universe we're used to." Turning a pawn into a queen could be part of the rules of relativity.

Such wildly speculative ideas may be closer to philosophy than to physics. But for now, quantum mechanics and general relativity—powerful, counterintuitive theories—are all we have to figure out the universe. "As soon as people start trying to bring quantum theory and general relativity into this, the first thing to say is that they really have no idea what they're doing," says Tim Maudlin, a philosopher of science at New York University. "It's not really rigorous mathematics. It's one piece of mathematics that sort of looks like general relativity and another little piece of mathematics that sort of looks like quantum theory, mixed together in some not entirely coherent way. But this is what people have to do because they honestly don't know how to go forward in a way that makes sense."

Will some future theory eliminate the possibility of time travel into the past? Or will the universe again turn out to be far stranger than we imagine? Physics has advanced tremendously since Einstein redefined our understanding of time. Time travel, which existed only in the realm of fiction for Wells, is now a proved reality, at least in one direction. Is it too hard to believe that some kind of symmetry exists in the universe, allowing us to travel backward in time? When I put the question to Gott, he replies with an anecdote:

"There's a story where Einstein was talking to a guy. The guy pulled a notebook out and scribbled something down. Einstein says, 'What's that?' The guy says, 'A notebook. Whenever I have a good idea, I write it down.'

Einstein says, ‘I’ve never had any need for a notebook; I’ve only had three good ideas.’ ”

Gott concludes: “I think we’re waiting for a new good idea.”

--Originally published: Scientific American 24(4); 26-31 (Winter 2015).

# Giant Bubbles of the Milky Way

by Douglas Finkbeiner, Meng Su and Dmitry Malyshew

On a clear night, away from city lights, you might see a beautiful structure arched across the sky: our home galaxy, the Milky Way. Since ancient times, humans have marveled at the dark dust clouds silhouetted against the milky background. Just four centuries ago Galileo pointed his telescope at the heavens and found that the "milk" seemingly splashed across the dark expanse was actually the blended light of countless stars.

The architecture of the Milky Way has now been revised again. We and our collaborators have discovered colossal structures that tower over the galactic center and extend tens of thousands of light-years into space. These luminous lobes have gone unnoticed for so long because they glow brightest in gamma rays, which cannot pass through our atmosphere. We needed an entirely new kind of telescope, orbiting in space, to see them.

We do not yet know what is creating these Fermi bubbles, as we have called them. But they appear to be driven by mysterious processes happening deep within the center of the Milky Way—a chaotic region where a supermassive black hole churns whirlpools of hot gas, while violent supernovae bloom like daffodils out of the rich soil of stellar nurseries.

Like many surprising discoveries, we found the Fermi bubbles serendipitously. Now we have begun to meticulously map their features. The giant bubbles of the Milky Way promise to reveal deep secrets about the structure and history of our galaxy.

**The Surprise Discovery**

The first hint that something was amiss in the inner galaxy came not from gamma rays but from microwaves. The year was 2003, and I (Finkbeiner)

was trying to better understand the origin of the universe using data from the Wilkinson Microwave Anisotropy Probe (WMAP), at the time the latest, greatest cosmology satellite. I was a postdoctoral fellow at Princeton University, studying how nearby interstellar dust obscured the signal from WMAP's intended target—microwaves from the dim afterglow of the big bang. The dust is interesting in its own right, but to a cosmologist it is like smudges on a window, a nuisance to be wiped away. To do that, we model the dust signal and subtract it from the data.

Because astronomers are forced to observe the cosmos from inside the Milky Way, I also had to subtract the microwave signals created by energetic particles (such as electrons) that fly through the galaxy. In 2003 astronomers already had a fairly sophisticated understanding of these signals, but something did not fit. I could model most of the galactic emission, but when I tried to subtract it from our data on the inner part of the galaxy, there was always something left over. I named this leftover signal the "microwave haze."

This mysterious signal coming from the center of the galaxy had no known explanation, but astronomers quickly came up with ideas. The most exciting possibility was that the haze was evidence of hidden dark matter. No one knows what dark matter is, only that it interacts with ordinary matter through gravity. Scientists expect that gravity will pull dark matter toward the center of the galaxy. In the dense cloud of dark matter in the Milky Way's core, dark matter particles will collide more often than elsewhere in the galaxy.

It is thought that dark matter may include both particles and anti-particles. If that is true, then colliding bits of dark matter and dark antimatter will annihilate each other and produce a cascade of intermediate particles. The cascade may ultimately end in the production of high-energy photons (gamma rays), plus a high-energy electron of ordinary matter and a positron—the electron's positively charged, antimatter counterpart.

We cannot see dark matter, but we should be able to see these particles it creates. As the electrons and positrons twist and turn through the tangle of magnetic fields at the galactic center, they should emit synchrotron radiation—the luminous exhaust of charged particles that are forced to make a turn.

The microwave haze we were seeing could have been an artifact of synchrotron radiation generated by dark matter. But how were we to tell for sure? The very same electrons that produce synchrotron microwaves should also be producing gamma rays through two distinct processes: deceleration by other charged particles, and collisions with photons.

If the microwave haze was being caused by high-energy electrons—perhaps as a consequence of dark matter annihilation—then we should also be able to find high-energy gamma rays by using the Fermi Gamma-ray Space Telescope, which launched in 2008. I had become a professor and was working in the summer of 2009 with Gregory Dobler, then my postdoctoral fellow, when data from the Fermi satellite was released to the public. We immediately rushed to make our first gamma-ray maps of the galaxy. After a few long days and nights, we found a hazy excess of gamma rays in the inner galaxy that appeared to match the microwave haze. We and our collaborators quickly submitted a paper arguing that the signals were related. We asserted that they are both probably caused by a high-energy population of electrons in the center of the galaxy, but we did not speculate about the source of the electrons.

The next shoe took a bit longer to drop. In October 2009 I was in my office remaking some figures in our first paper with newly released Fermi data. I had noticed that the original gamma-ray data showed faint edges—clear borders where the signal dropped off precipitously. In astronomy, sharp features usually come from transient events. For example, a supernova may send out a shock wave that appears as a distinct edge in our telescopes. In time, sharp features tend to smooth out and fade away.

If dark matter were causing the gamma-ray signal, then the drop-off should have been smooth—fading gently farther away from the galactic center—because dark matter annihilation would have been going on for billions of years. Any sharp edges would have dissipated long ago.

In the first batch of Fermi data, the edges had looked so ratty that we just chalked them up to noise in the signal and ignored them. Now they were appearing in the new data again, and I started to wonder. I showed them to my then graduate students Meng Su and Tracy Slatyer, who agreed that they were real. Then Su really jumped in and started to work—I think almost continuously without sleep—on deriving the exact shape of the edges.

Within a matter of days we totally changed our opinion about what was in the data. Dark matter was out. Bubbles were in. In May 2010 Su, Slatyer and I submitted a paper to the *Astrophysical Journal* describing the structures and naming them "Fermi bubbles" in honor of the Fermi telescope.

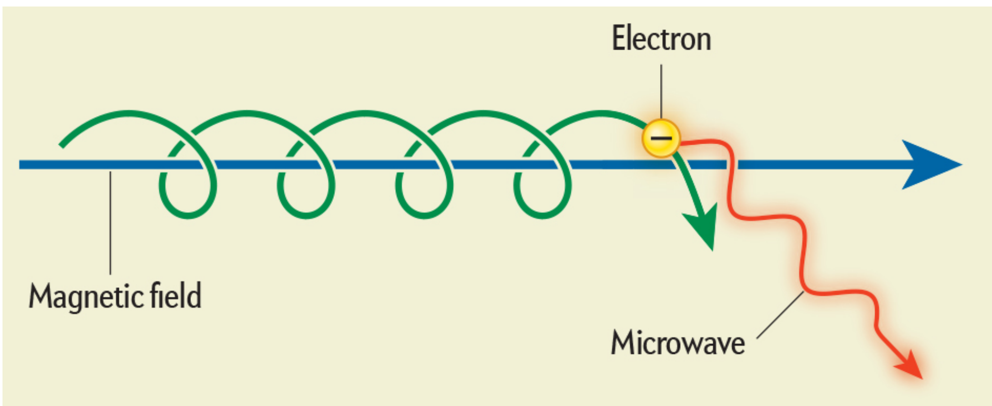
---

## A Field Guide to Cosmic Radiation

Each of the three radiation-producing processes below came into play as the authors teased out the story of the Milky Way's Fermi bubbles. The radiation originates with electrons accelerated to high energies near the center of the galaxy.

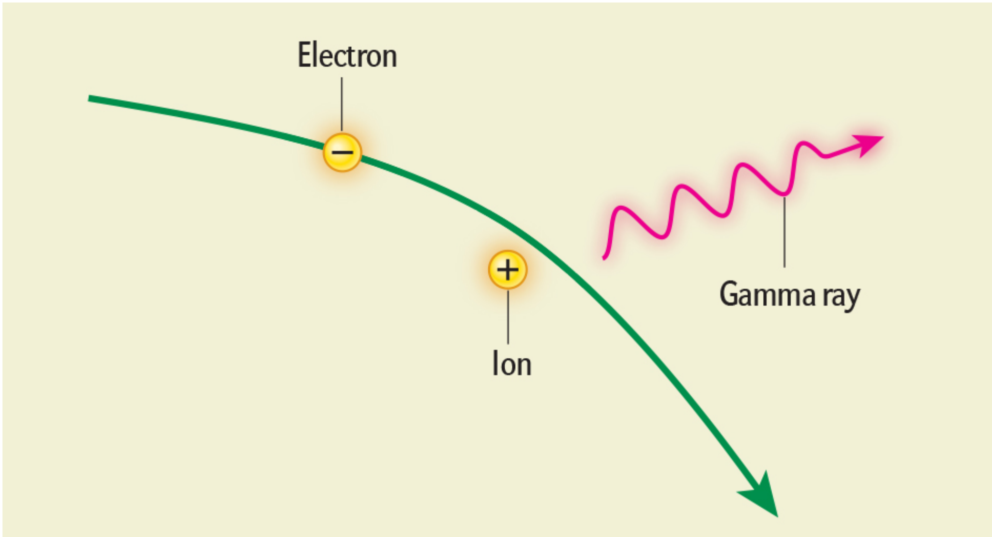
**Synchrotron**

When charged particles such as electrons change direction, they emit radiation. In the central Milky Way, strong magnetic fields spin electrons in circles, generating so-called synchrotron radiation. Galactic synchrotron radiation comes mainly in the form of microwaves.



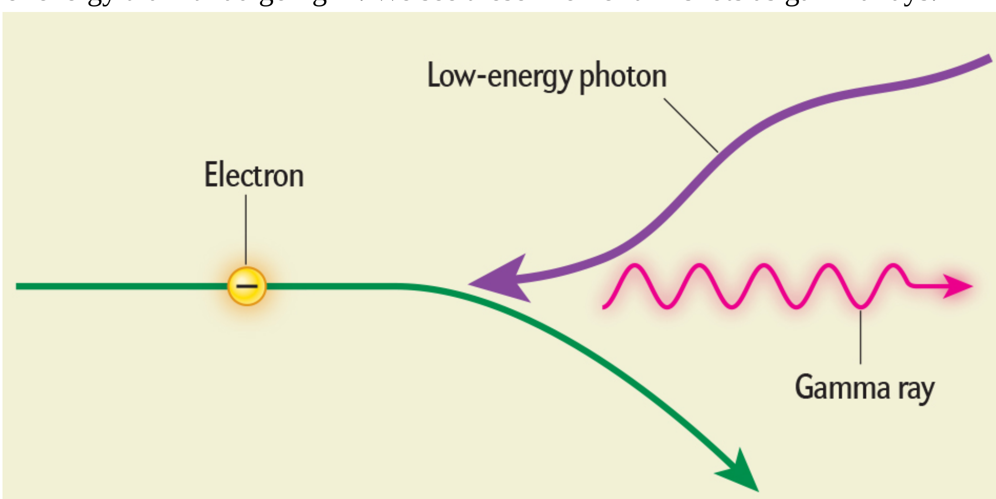

**Bremsstrahlung**

If an energetic electron passes by another charged particle, the electron will often slow down, losing energy in the process. This lost energy becomes a photon—a particle of light—which shoots out of the electron as bremsstrahlung radiation. For high-energy electrons, these photons are gamma rays.




**Inverse Compton**

An electron flying through the galaxy can also smack head-on into a photon. Much like a baseball encountering a well-swung bat, the photon shoots out of the collision with much higher energy than it had going in. We see these "home run" shots as gamma rays.




*Graphic by Greg Maxson*

---

## Bubble Makers

Even though nobody had expected to find bubbles made of high-energy electrons and atomic nuclei (known as cosmic rays) jutting tens of thousands of light-years above the Milky Way, perhaps it should not have been that shocking.

Many other galaxies have bubbles, too. We can see them in x-rays and radio waves. If we had better gamma-ray telescopes, we would probably find them shining in gamma rays as well.

We understand the processes that create the bubbles in many of these other galaxies. In some cases, the bubbles trace their origin to a gigantic black hole—often having the mass of billions of suns—that anchors the galaxy's center. As material from the galaxy falls toward the black hole, it begins to spin like the water draining out of a bathtub. This whirlpool of hot gas and dust creates intense magnetic fields that power jets of radiation and cosmic-ray particles that may inflate the bubbles.

We know that the Milky Way galaxy also has a supermassive black hole at its center, but we have never observed a strong jet of intense radiation streaming out of its core. (If a jet exists, it is not pointed our way, and thank goodness for that.) So we do not have direct evidence that this process is inflating the Fermi bubbles.

On the other hand, a large gas cloud—the Magellanic stream—sits high above the galactic center. If a jet of radiation were pointed there, it would temporarily strip electrons free from atoms in the cloud. As the electrons and ions came back together, the recombination would produce radiation.

In 2013 astronomers found exactly this. Perhaps there was an intense episode of accretion onto the Milky Way's central black hole several million years ago—a high-speed whirlpool of hot, infalling matter that generated high-energy jets and ultraviolet radiation. The radiation in turn would have knocked the electrons around in the Magellanic stream. This event could have also created the Fermi bubbles.

Alternatively, some galaxies have bubbles that are by-products of intense star formation near their centers. In a stellar nursery, stars form in many different sizes. The more massive a star is, the faster it burns its nuclear fuel. When the fuel runs low, the star's core collapses and releases an enormous amount of energy that rips off the outer layers of the star in a supernova explosion, leaving a neutron star or black hole behind. Collectively, these supernovae create a wind of particles that can inflate bubbles around a galactic center.

We know that the center of the Milky Way has also been a region of intense star formation. Several thousand stars around the central black hole are only about six million years old—mere toddlers in cosmic time. Yet if extremely massive stars also formed in this same stellar nursery, six million

years would be long enough for them to have already exploded as supernovae. These supernovae would have driven a wind of hot gas out from the galactic center—a wind that might have been powerful enough to inflate the bubbles.

**Illuminating the History of the Galaxy**

The story of the Fermi bubbles is wound tightly with the history and evolution of the Milky Way. The bubbles, which recent observations suggest formed about 2.5 million to four million years ago , may shed light on how the black hole at the galaxy's center formed and evolved. The bubbles can also teach us about the physics of how black holes pull in nearby matter and how high-energy cosmic rays interact with interstellar gas. Although structures like the Fermi bubbles exist in other distant galaxies, having an example in the Milky Way lets us study these systems up close.

To this end, we are trying to observe the bubbles using the entire electromagnetic spectrum. One of the most amazing things about the bubbles is that they are so large and luminous in gamma rays yet nearly invisible at other frequencies. New data from the Planck spacecraft, which has mapped microwave radiation across the entire sky, are providing important clues. The Dark Matter Particle Explorer satellite, scheduled for launch in late 2015, will map gamma rays of higher energy than we have seen so far with the Fermi Large Area Telescope.

We are also attempting to map the bubbles in x-rays, although we are limited by current technology. The bubbles are giant structures that tower over the galaxy, but nearly all x-ray satellites currently in orbit have a narrow field of view. The challenge is akin to mapping a mountain range while peering through a soda straw. We look forward to the launch in 2017 of the Spectrum- Roentgen-Gamma satellite, which is designed to produce a new survey of the sky in medium-energy x-rays.

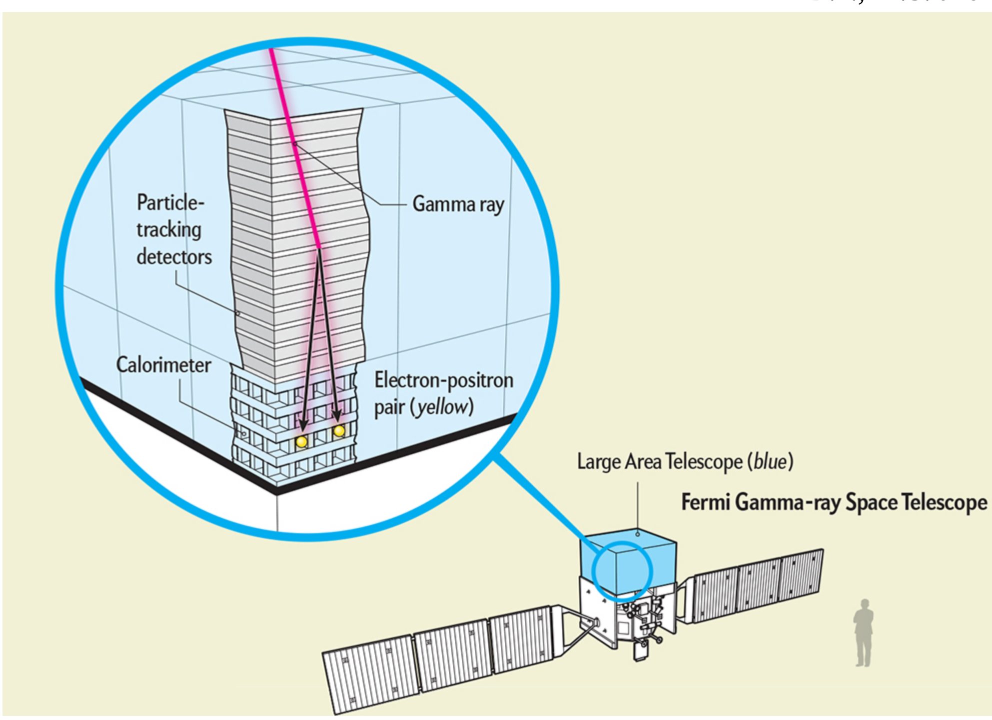
It took three centuries from Galileo's discovery that the Milky Way is made of stars for astronomers to realize that our galaxy is just one of many billions of galaxies spread throughout the cosmos. With any luck, we will come to understand the true significance of the Fermi bubbles in less time than that.

# A Gamma-Ray Eye

Earth's atmosphere blocks gamma rays, which have energies billions of times that of visible light, so one way astronomers measure them is by getting above the atmosphere. The Fermi Gamma-ray Space Telescope is the most powerful gamma-ray observatory ever launched. It contains two main instruments: a burst monitor (not shown) that surveys the entire sky for evidence of transient gamma-ray bursts and the Large Area Telescope (LAT), which is the most sensitive and highest-resolution gamma-ray detector ever launched.

The LAT is radically different from any optical telescope: it has no mirrors, no lenses and no focal plane. Instead it operates more like a particle physics experiment. Each incoming gamma ray recoils off an atomic nucleus in the telescope and transforms into an electron and its antimatter counterpart, a positron. These particles are then tracked through onboard detectors and a calorimeter, which measures energy. Further data analysis on the ground filters out background noise and reveals the direction and energy of the original gamma ray. Most telescopes can see only a tiny fraction of the sky at a time, and astronomers spend a great deal of effort deciding which parts of the sky to observe. Competition for telescope time is fierce, and it is generally not feasible to observe a large swath of sky where nothing interesting is expected. In stark contrast, Fermi has a field of view covering a fifth of the sky, which allows it to observe the breadth of the sky every three hours. This full-sky coverage gives astronomers the chance to find large, faint surprises like the Fermi bubbles.

— D.F., M.S. and D.M.




*Graphic by Greg Maxson*

---

--Originally published: Scientific American 24(4); 32-37 (Winter 2015).

# SECTION 2

# Extreme Machines

# Neutrinos at the Ends of the Earth

by Francis Halzen

One of the most ambitious and extreme experiments on earth opened at the South Pole in 2010. IceCube, a giant particle detector buried in the polar ice, captures elusive, high-energy species of neutrinos—fundamental particles that fly straight through almost everything they touch. The project, for which I am the principal investigator, aims to use neutrinos to study distant cosmic phenomena—particularly the mysterious, violent processes thought to produce the charged particles known as cosmic rays, which continually bombard Earth.

We expect IceCube to catch superenergetic neutrinos only rarely. The particles have almost no mass and no electric charge (which is why they seldom react with other particles), and they travel at nearly the speed of light. Once they arrive on Earth, most neutrinos do not stop to linger. They zip right through thousands of kilometers of rock—even through the planet's solid iron core—and continue on their way.

Because of these difficulties, we were not surprised that data taken during the first few years, while the detector was still under construction, turned up nothing extraordinary. But in 2012 that changed.

One day, during a routine conference call for team members, our screen lit up with patterns we had never seen before. The signals reflected two neutrinos carrying more than 1,000 times the energy of the most energetic neutrino ever produced by an accelerator on Earth. These wild neutrinos had almost a billion times the energy of the neutrinos that the sun regularly spits out. Clearly, they had come from some spectacularly energetic process occurring far from our planet.

Exhilaration spread through the room as we realized we were looking at something game changing. Capturing the whimsy of the moment, one of our graduate students nicknamed the two particles “Bert” and “Ernie,” after the Sesame Street characters (fun names are easier to keep straight than the long strings of numbers we usually assign to neutrino events).

It took us another year and a totally redesigned analysis of the same data to satisfy ourselves that these were indeed what they seemed: the first pixels of the first pictures of the distant neutrino universe. Since then, we have found 54 high-energy neutrinos in total—many of them given Muppet names, including one dubbed “Big Bird” that had an energy twice that of Ernie or Bert.

We are now trying to identify where in the sky these high-energy neutrinos came from and how they originated. Their suspected sources are extreme cosmic events such as supernovae and other stellar explosions called gamma-ray bursts—two phenomena rumored to give rise to cosmic rays. If we can definitively trace the high-energy neutrinos to these likely sources of cosmic rays, we will open a new frontier in our understanding of the physics behind the extraordinarily dramatic events that are thought to produce them.

**Powerful Particles**

The earth is bathed in cosmic rays arriving from outer space—IceCube detects 275 million of them every day. A cosmic ray is a charged particle, most often a proton, that has extremely high velocity and thus carries a lot of energy. More than a century after their discovery, the processes that create cosmic rays are still largely unknown. When they arrive at Earth, we cannot simply trace their trajectory backward to deduce where they came from, because they swerve about during their journey, deflected by galactic and intergalactic magnetic fields.

Luckily, however, theory suggests that cosmic rays also interact at their birthplaces with photons to produce neutrinos. And neutrinos do point back to where they started—they shun other matter so thoroughly that almost nothing can divert them from their path. So although we cannot pinpoint the origins of cosmic rays from the rays themselves, we can infer their

birthplaces by analyzing the highly energetic neutrinos they presumably produced in their youth.

Of course, astronomers have some ideas about how cosmic rays are born, but we need data to help us confirm or discard those possibilities. It is thought likely, for example, that massive stars give off cosmic rays in their death throes. At the end of a large star’s life, when its nuclear core can no longer support its mass, it collapses into a dense object called a neutron star or into an even denser black hole, from which nothing escapes. The collapse set off an incredibly bright explosion: a supernova. But it also converts large amounts of gravitational potential energy into kinetic energy—thrust for the acceleration of particles, presumably through shock waves.

Supernova remnants were proposed as a likely source of cosmic rays as early as 1934 by astronomers Walter Baade and Fritz Zwicky. Yet, 80 years later, astrophysicists still cannot agree whether the hypothesis is correct. It does strike many as plausible. About three stars in the Milky Way go supernova every century. And each supernova converts a reasonable fraction of the star’s mass into fuel for particle acceleration. But we need more evidence to be certain that supernovae alone account for the steady flow of cosmic rays seen in the galaxy.

Extragalactic cosmic rays, which originate from beyond the Milky Way galaxy, generally pack even more energy than do cosmic rays coming from nearby. There must be some source more energetic than supernovae that creates them.

Gamma-ray bursts are a prime suspect. Even brighter than regular supernovae, gamma-ray bursts are somewhat mysterious but probably occur when stars of very high mass collapse under certain extreme conditions.

But there are competing ideas about the engines of extragalactic cosmic rays. It is possible they originate near active galactic nuclei—supermassive black holes that sit at the centers of their galaxies and consume matter voraciously. As clouds of dust or gas get pulled in toward such a black hole, some particles could be deflected outward at ultrahigh speeds and eventually reach us as cosmic rays.

## To Catch a Neutrino

Spotting the neutrinos that point to the origins of cosmic rays is a game of long odds. To have any hope of seeing dozens or hundreds of these elusive particles within a few years, IceCube had to be gigantic. The experiment is now monitoring a cube of 100,000-year-old Antarctic ice that is 1.5 kilometers below the surface of the South Pole. The cube measures a full kilometer on each side.

Deep, old Antarctic ice is a perfect natural neutrino detector because it is pure, ultratransparent, and shielded from sunlight. When a neutrino does occasionally interact with atoms in the ice, a shower of charged particles radiate blue light, called Cherenkov radiation, that can travel hundreds of meters through the ice. IceCube's 5,160 optical sensors, encased in glass spheres and spaced evenly throughout the cube, are able to pick up these faint flashes.

The sensors chart, in exquisite detail, the light pool produced by the nuclear debris created when a single neutrino hits. This pattern reveals the neutrino's type (or "flavor"), energy and arrival direction. The energies of Ernie and Bert and the others that we have seen so far are about a peta–electron volt (PeV), or $10^{15}$ eV; Ernie and Bert were 1.07 PeV and 1.24 PeV, respectively. For comparison, the particle beams at the Large Hadron Collider at CERN near Geneva, the world's most powerful particle accelerator, are in the tera–electron volt (TeV), or $10^{12}$ eV, range, about a thousandth as energetic. Big Bird, Ernie and Bert are the most energetic neutrinos ever seen , by a wide margin. Each of them cast a light pool of roughly 100,000 photons spread over about six city blocks.

The PeV energies of these two neutrinos tell us something important: they must be part of some cosmic signal. Their energies are just too large to have been produced nearby.

Local neutrinos, in contrast, are a dime a dozen. Every six minutes, on average, IceCube detects a neutrino made in Earth's atmosphere when cosmic rays smack into hydrogen and oxygen nuclei. But neutrinos made in our own backyard tell us nothing about the nature of cosmic rays or other astrophysical phenomena.

Our analysis of the data generated by IceCube thus starts by screening out these distractions. From past experience, we know what kinds of light

patterns garden-variety neutrinos make. We ignore those.

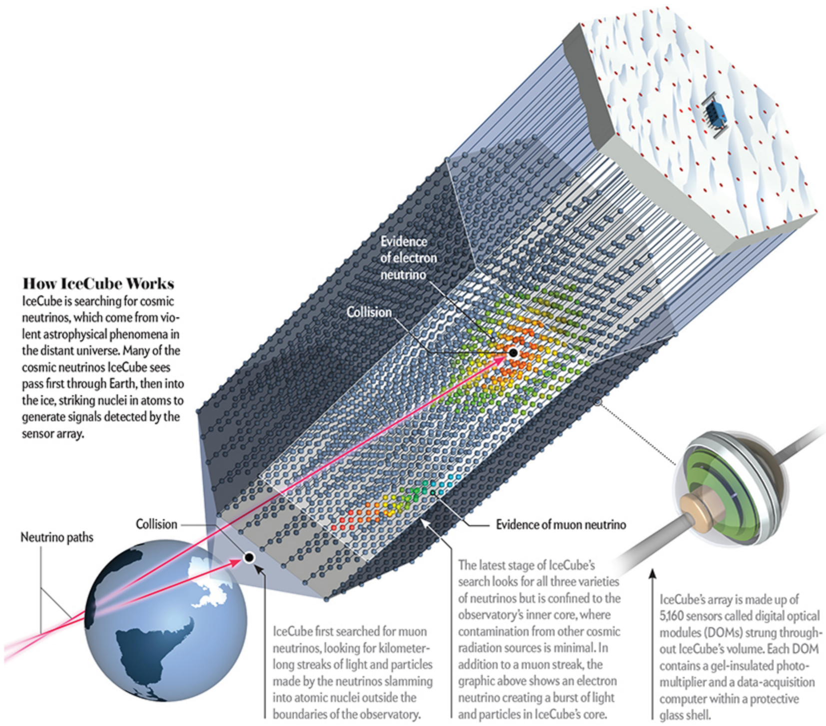
Any flashes that remain and correspond to PeV-energy neutrinos must come from the distant cosmos. They very well could have reached us from the same sources as cosmic rays. But there are other plausible, even more exotic explanations for these particles.

Some have suggested that they may be signatures of dark matter—the invisible material that seems to make up more than 80 percent of all matter in the universe. If dark matter consists of very heavy particles having an average lifetime longer than the current age of the universe, then the occasional decay of a dark matter particle might produce the PeV-energy neutrinos that we observe.

---

### Ghost Hunting

Neutrinos flit like ghosts through the universe. They come in three varieties—muon, tau and electron—but all three are devoid of electric charge and are so light and fast that they scarcely interact with other particles. Trillions whiz harmlessly through your body every second, and most pass through Earth itself as if it were empty space. Very rarely, however, individual neutrinos can strike the nucleus of an atom, creating cascades of subatomic particles and brief flashes of light that allow researchers to study these elusive particles.



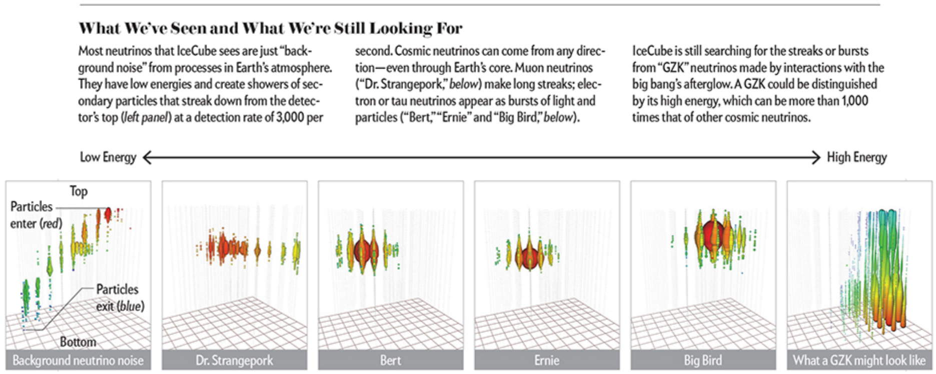



*Illustration by George Retseck*
*Source: Courtesy of IceCube Collaboration (bottom six panels)*

## Counting Neutrinos

Before IceCube discovered PeV neutrinos, the search for cosmic neutrinos had focused almost exclusively on muon neutrinos and paid little regard for the other two flavors, known as electron and tau neutrinos. That's not because muon neutrinos are thought to be the most common variety of cosmic neutrino reaching us—that flavor just happens to be easier to spot in our detector. When muon neutrinos slam into atoms, they launch kilometer-long light trails.

We originally optimized IceCube to pick up muon-neutrino trails that extend into the detector volume, even if the trails originate outside the cube. In effect, this technique allows us to expand our observation volume. But it comes with a trade-off. The approach also increases the risk that particles other than cosmic neutrinos will contaminate the result. Screening out background noise thus becomes harder.

So we also ran a second analysis that homed in on just the inner half of the IceCube detector, a strategy that leaves less room for contamination to get in. The great advantage of confining the search in this way is that the detector can then measure the full energy that each neutrino deposits in the ice to within 10 to 15 percent accuracy. That is a big improvement on the measurements we can make of neutrinos that interact outside the detector.

This second, more tightly constrained search hunted specifically for a class of extremely high energy neutrinos called Greisen-Zatsepin-Kuzmin (GZK) neutrinos. Theorists predict that when cosmic rays interact with photons from the cosmic microwave background, which was left over from

the big bang, neutrinos could emerge having energies as high as exa–electron volts (EeV)—roughly $10^{18}$eV.

Neither we nor anyone else have yet found GZK neutrinos. But IceCube's search for them has turned up plenty of cosmic neutrinos in all three flavors.

Since the discovery of Ernie and Bert, both our search strategies have succeeded in detecting cosmic neutrinos. Our first two years' worth of data revealed 28 neutrinos with energies between 30 and 1,200 TeV, including Ernie and Bert. This number is more than four standard deviations above what we would expect purely from the atmospheric background. Looked at another way, the probability is greater than 99.9999 percent that these particles truly came from deep space.

When we later added an additional two years of data, we brought the tally to 54 cosmic neutrinos. The statistical significance of the signal then climbed to well over five standard deviations, the conventional threshold for a "discovery."

Exactly where in the universe do all these neutrinos point back to? We have not yet collected a large enough sample of events to answer that question conclusively. The origins do not seem to be restricted to our galaxy—the sky map indicating their arrival directions shows only marginal evidence for an overlap with the plane of the Milky Way. Most of the cosmic neutrinos are almost certainly of extragalactic origin.

There appears, however, to be a somewhat higher than average number of neutrinos coming from the center of the Milky Way. Bert is part of that cluster; it points back to within one degree of the galactic center. We cannot say for sure why this region is spewing out relatively high numbers of neutrinos, but we know the galactic center is packed with supernova remnants, as well as a giant black hole. Any of these are likely candidates for the neutrinos' source.

Our goal is to continually refine the map of where cosmic neutrinos originate as we steadily collect more muon-flavored neutrinos reaching us through Earth. The kilometer-length light trails they shed allow us to reconstruct their trajectories with better than 0.5-degree resolution. Over time, the accumulating data will reveal a highly detailed map of the sky in

high-energy neutrinos—and therefore cos mic rays. Astronomers will be analyzing the map to find overlaps with known celestial objects that could be sources, such as gamma-ray bursts or bright galaxies that host supermassive black holes and active galactic nuclei.

IceCube is just beginning to scratch the surface of what it can discover. The experiment is built to run for 20 years—maybe more. In the meantime, we are looking toward its sequel. Our team is proposing to eventually build an expanded detector using roughly 10 cubic kilometers of ice—about 10 times the volume of IceCube. This larger instrument should collect thousands more cosmic neutrinos, enough to determine once and for all what distant powerhouses are creating them and their cosmic-ray cousins.

--Originally published: Scientific American 24(4); 38-45 (Winter 2015).

# Seeing in the Dark

by Joshua Frieman

Nearly 100 years ago Edwin Hubble discovered that the universe is expanding: almost all galaxies are speeding away from our own Milky Way, and faraway galaxies are receding faster. That discovery was profound, but it was followed, in 1998, by an even more startling realization: the expansion is accelerating. For most of the 20th century, scientists had expected that over time gravity would pull galaxies toward one another, slowing the expansion. In stead two teams of astronomers studying supernovae—exploding stars that serve as markers for measuring cosmic distances—found that the growth of the universe is actually speeding up. This remarkable discovery, since confirmed by other observations, was awarded the Nobel Prize in Physics in 2011. But *why* is the universe accelerating? This enigma is one of the biggest unsolved mysteries in all of science.

To explain it, cosmologists have come up with two alternative ideas, either of which would revolutionize our understanding of the laws of nature. One is that Isaac Newton (and more recently, Albert Einstein) did not have the last word on gravity: although gravity is attractive on Earth and in the solar system, perhaps it acts differently, becoming a repulsive force, when it comes to the vast distances of intergalactic space. Maybe we need to modify the theory of how gravity operates on cosmic scales.

The other idea is that the universe is filled with some unseen stuff—now called dark energy—that counteracts the force of gravity, making objects repel instead of attract one another. Cosmological measurements indicate that dark energy, if it exists, currently makes up about 70 percent of the universe by mass or energy (mass and energy are equivalent, as Einstein showed with his equation $E = mc^2$). Dark matter (no relation to dark

energy), an invisible form of matter, makes up about 25 percent, and normal matter—things made of atoms, including stars, planets and people—contributes only about 5 percent. This picture has garnered more attention than the notion that gravity works differently on large scales because it neatly explains the formation of galaxies and larger structures in the universe and is consistent with all measurements to date.

But how can we know for sure if dark energy is to blame for cosmic acceleration? And if it *is* dark energy, what is the nature of this stuff? We recently launched an ambitious project called the Dark Energy Survey (DES) to better understand why the universe seems to be ripping apart.

The survey should provide answers by gathering a thorough record of the 14-billion-year history of cosmic expansion and the rate of growth of large-scale structure—the vast conglomerations of galaxies spread across the universe—with unprecedented precision. By studying how structures grouped together over time, we hope to distinguish among the various possibilities for why they are pulling apart now.

My colleagues and I at Fermi National Accelerator Laboratory and the University of Chicago, along with 300 other physicists and astronomers from 25 institutions in the U.S., Spain, the U.K., Brazil, Germany and Switzerland, make up the DES collaboration and have worked together to build, operate and analyze data from the Dark Energy Camera, the hardware heart of our project.

In 2012 we mounted this camera on a four-meter-diameter telescope at Cerro Tololo Inter-American Observatory, a U.S. facility high in the Andes Mountains of northern Chile. It took its first snapshots of the night sky that September and was commissioned over the following months. On August 31, 2013, the DES officially began surveying a large swath of the southern sky. The survey, now in its third season, will run from August to February every year for five years and ultimately produce a deep, high-resolution map of about 200 million galaxies spread over one eighth of the sky as well as a catalog of stellar explosions that can be used to track cosmic expansion. The survey has already collected a wealth of data that is being analyzed and on the way toward unlocking the secret to the universe's expansion.

## Competing Hypothesis

Fortunately for scientists, the same evidence that should distinguish between the modified gravity and dark energy hypotheses of acceleration should help clarify what dark energy is, if it exists. The survey will test two main ideas about dark energy . The simplest explanation for it may seem counterintuitive: that it is the energy of empty space. Suppose you took a box and emptied it of all matter—all the atoms, radiation, dark matter, and so on—and nothing could penetrate its walls. The inside of the box would be a perfect vacuum. According to classical physics, the vacuum—empty space—has no energy. But quantum theory says that even empty space carries energy. Physicists think of this energy as coming from "virtual" particles: at any time a particle and its antiparticle can appear spontaneously for a brief instant, then annihilate each other and disappear back into the vacuum. Virtual particles carry energy in exactly the form that would be needed to constitute dark energy and cause the expansion of the universe to speed up.

The only difficulty with this notion is that quantum physics predicts that the amount of vacuum energy in space should be 120 orders of magnitude ($10^{120}$) larger than what it seems to be if it is responsible for dark energy. If you are working on a math problem, it is hard to make an error that big. Partly as a result of this discrepancy, cosmologists have proposed other explanations for dark energy aside from vacuum energy.

One idea—the second notion being tested by the survey—is that dark energy takes the form of a so far undetected particle that could be a distant cousin of the recently discovered Higgs boson: it would have some of the properties of the Higgs particle but would be 44 orders of magnitude lighter. This possibility is sometimes dubbed "quintessence." One can think of such a particle as acting like a ball rolling down a hill at each point in space. The rolling ball carries both kinetic energy (because of its motion) and potential energy (because of the height of the hill it is rolling down); the higher an object is, the greater its potential energy is. As it rolls down, its potential energy declines, and its kinetic energy rises. If the quintessence particle is extremely light, with a mass less than about $10^{-33}$ electron volt (by comparison, the mass of the electron is 511,000 electron volts), then it would be rolling very slowly today, with relatively little kinetic versus

potential energy. In that case, its effect on cosmic expansion would be similar but not identical to that of vacuum energy and would lead to acceleration, although most versions of quintessence predict that the acceleration would begin later in cosmic history than if vacuum energy were the culprit.

## Four Probes

To distinguish among the possible causes of cosmic acceleration, the Dark Energy Survey—which is funded by the U.S. Department of Energy and the National Science Foundation, with additional support from the participating institutions and foreign funding agencies—is investigating four phenomena that are particularly sensitive to whatever is pulling the universe apart. And because each involves a different observable quantity, the four probes will not all be affected by the same measurement errors.

These four phenomena are supernovae, signatures of primordial sound waves, gravitational lensing (the bending of light by gravity) and galaxy clusters. Collectively they tell us how fast the universe has expanded and how much matter has clumped together to form large-scale structures at different epochs of cosmic history. At early times, up to about several billion years after the big bang, gravity fought against the expansion and enabled large-scale structures to form. But when the universe was around seven billion years old, matter became dilute enough that whatever was causing accelerated expansion—be it dark energy or modified gravity—became dominant over gravity and sped up the expansion, gradually shutting down the further formation of large structures. Vacuum energy, quintessence and modified gravity would each leave unique signatures in the history of the cosmic expansion rate and in the pattern of structure growth, imprints that we can tease out through these four probes.

---

### Cosmic Acceleration: Four Approaches

The universe is expanding at an ever faster rate, and the Dark Energy Survey (DES) aims to find out why by observing four different signals. The left two measure cosmic distances to see how large the universe was and how fast it was expanding at various epochs. The right two map the clumpiness of matter throughout space to gauge the competition between gravity and the force behind the universe's acceleration. Certain patterns of expansion and clumping over time would suggest that the universe's acceleration stems from one or another form of dark energy—a hypothesized constituent of space that counteracts the force of gravity. Other patterns, meanwhile, could indicate that to explain the acceleration, scientists must rewrite the laws of gravity for cosmic scales.

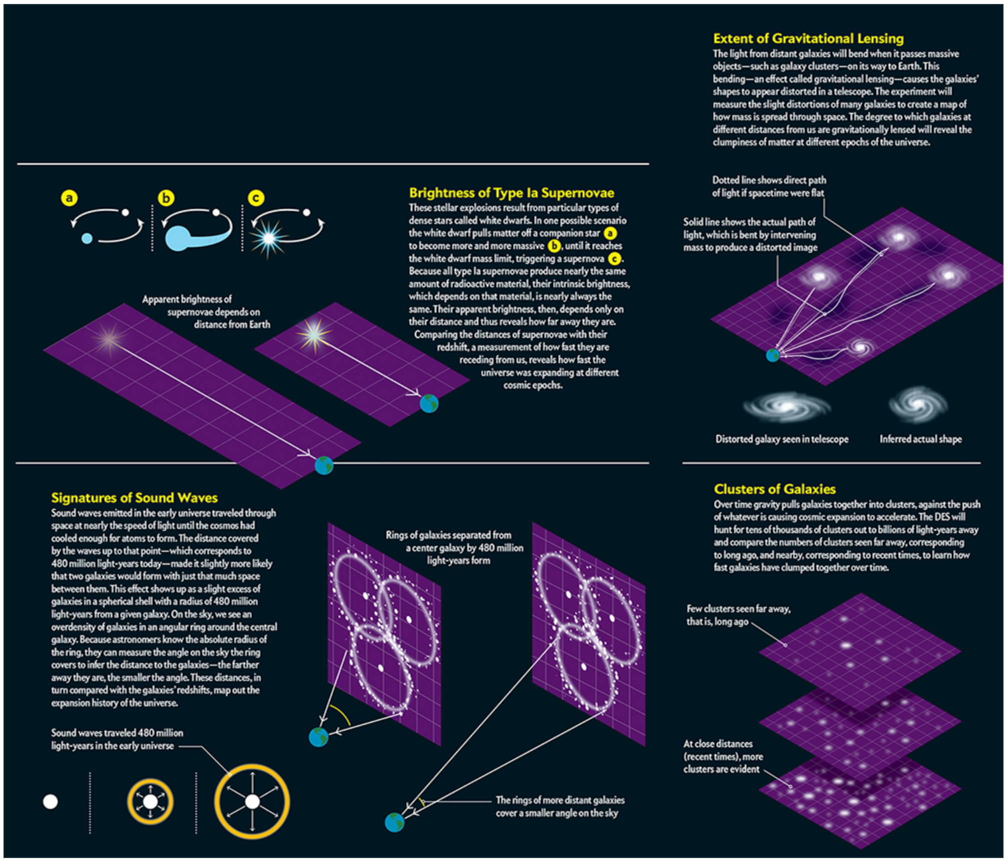



*Illustration by Nigel Hawtin*

---

## Supernovae

A type 1A supernova is a stellar explosion that results when a small, dense object called a white dwarf star reaches a certain mass limit. These supernovae all reach a peak brightness that is nearly the same. Any differences in how bright they appear to us stem solely from their distance: those that look dimmer are farther away. This feature makes them so-called standard candles, or good cosmic yardsticks. We know, for instance, that a type Ia supernova 100 times as faint as another is 10 times farther away.

The DES will observe the same few patches of sky every few nights to measure accurate distances to a few thousand type Ia supernovae in the nearby and distant universe, nearly 100 times as many as were used in the 1998 discovery of cosmic acceleration. We are also using other telescopes to measure how much the light from these supernovae is shifted toward the red end of the visible spectrum. This redshift occurs for any object speeding away from an observer and tells us how much the wavelength of the light has been stretched by the expansion of the cosmos between the time it was

emitted and now. The redshifts of the distant supernovae directly reveal the relative size of the universe then versus today. Taken together with the standard candle distance measurements to the same objects, the DES will be able to reconstruct the past 10 billion years of the expansion history of the universe with great precision.

Such a measurement can distinguish among different theories of cosmic acceleration because each would have produced a slightly different expansion history. If quintessence is at work, for example, accelerated expansion would probably have started somewhat later than in the vacuum energy scenario and built up more gradually. Thus, supernovae of a given redshift will appear brighter—will be closer—if the universe contains a Higgs-like quintessence particle than they will if vacuum energy is driving expansion. And if gravity works differently than we think, the pattern among distant supernovae will again differ, although the details vary depending on the specific modifications investigators have proposed to the classic idea of how gravity works.

Very high precision in these measurements is necessary to distinguish among the different models because their predictions diverge only slightly. Therefore, we wish to know the distance versus redshift relation to roughly single-percent-level accuracy—a feat that the Dark Energy Camera, for the first time, should be able to manage.

**Signatures of Primordial Sound Waves**

The DES will also use a relic from the beginning of the universe to study its expansion history. In the early universe, gravity was pulling matter together while the outward pressure of the electromagnetic radiation (light) in the cosmos resisted such compression. This competition created a series of sound waves. A few hundred thousand years after the big bang, when ordinary matter had cooled sufficiently from its initial hot state to transition from an ionized gas into atoms, the atoms and radiation went their sep arate ways (they effectively stopped interacting with one another), and this competition ceased. The distance traveled by the sound waves up to that point, which today corresponds to a scale of about 480 million light-years, ended up imprinted in the spatial distribution of galaxies as a slight tendency for pairs of galaxies to be separated by this distance compared with other distances.

This baryon acoustic oscillation (BAO) scale provides a standard ruler for measuring cosmic distances and the expansion history. That is, if you know the physical size of a ruler (the 480-million light-year spread of many galaxies from one another) and can measure how big it appears (the angle of separation between those galaxies on the sky), then you can tell how far away it is. The DES will measure this BAO feature for about 200 million galaxies, enabling us to chart their distance versus their redshift as we do for supernovae. Galaxies of the same redshift would be closer if quintessence caused the universe to start accelerating later than if vacuum energy was responsible and acceleration began earlier. If there is no dark energy, we expect that the relation between distance and redshift will look different from either of those scenarios, although the particulars will again depend on how exactly gravity is altered.

### Gravitational Lensing

This method focuses on a feature of light predicted by Einstein's general theory of relativity. The paths of light rays, as they travel to Earth from distant galaxies, get bent by the gravitational field of the matter they pass. This bending leads to a distortion of the images of these galaxies, an effect known as gravitational lensing. When the bending effect is large, the resulting images can be dramatic: distant galaxies may appear as thin, very extended arcs of light, and one may even see multiple images of the same galaxy. The light rays from most galaxies, however, are bent only slightly, leading to very small distortions in their shapes that are not discernible by eye: this is the regime of *weak* gravitational lensing.

Light rays from equally distant galaxies near to one another on the sky get bent by nearly the same amount because they travel through roughly the same intervening matter. By measuring the shapes of many galaxies in a small patch of sky, we can infer how much the images have been distorted and thus the clumpiness of the intervening matter, even though each galaxy image is distorted only slightly. Repeating this measurement for galaxies in different parts of the sky thus reveals the general clumpiness of matter in the universe. The evolution of this clumpiness, because it reflects the competition between gravity and dark energy and is sensitive to any modification of gravity, can help tell us what is causing the universe to accelerate.

The DES will measure the shapes of 200 million galaxies to see this effect, covering over 20 times more galaxies and a greater area of sky than previous weak lensing studies. By making extremely precise measurements of the shapes of these galaxies across the sky at different distances from Earth, we can create the most precise map yet of the distribution of matter at various removes—that is, at various cosmic time periods because the farther something is from Earth, the longer its light takes to reach us.

The map will differ depending on what is pulling the universe apart. The effects of quintessence in hindering the growth of large-scale structure, for example, probably set in during an earlier cosmic epoch than those of vacuum energy. Because we know from measurements how clumpy the universe is today, if quintessence were at work, we would expect to see more clumpiness when the universe was younger than in the case of vacuum energy. That prediction may sound counterintuitive because dark energy would hinder clumps from forming, but for the universe to have its current structure after billions of years of expansion, it would have to have been relatively clumpy early on. If there were no dark energy, modified gravity would have led to yet a different pattern of clumpiness throughout time—although whether the clumpiness would be relatively more or less at early epochs would differ for different formulations of the laws of gravity.

**Galaxy Clusters**

Finally the DES will also hunt for clusters of galaxies to trace cosmic clumpiness over time. Clusters, having masses of up to $10^{15}$ (1,000 trillion) times the mass of the sun, are the largest gravitationally bound objects in the universe, and they form against the pull of either dark energy or modified gravity. Unlike previous cosmological cluster surveys, which have been limited to smaller areas of sky, the DES aims to discover tens of thousands of clusters out to billions of light-years away.

Scientists will then compare the number of clusters they see close to Earth—corresponding to recent times—and far away in the past. Similarly to the effects on matter’s clumpiness as shown by weak gravitational lensing, we expect to see more clusters in the early universe if quintessence is at work than if vacuum energy is shaping the cosmos (all other things being equal), and we would see a different and more complicated trend altogether if gravity behaves unusually.

## State-of-the-Art Instrument

The secret weapon for our project is the most powerful camera ever made for looking into this question. The Dark Energy Camera, mounted on the Victor M. Blanco Telescope, is designed to survey numerous objects, including galaxies, clusters and super novae, in the shortest possible time. The ultrasensitive, 570-megapixel camera has a very large field of view, enabled by five large lenses, best for taking in large swaths of the universe at a go.

Since its official start in August 2013, the survey has covered nearly 5,000 square degrees of sky, obtaining color images of about 100 million galaxies. The supernova survey has discovered more than 1,000 type Ia supernovae so far. We are now analyzing these data to extract information about the supernovae's distances to compare with redshift. We are also measuring galaxy shapes to infer the weak lensing signal, identifying distant clusters of galaxies and measuring their properties, and measuring the spatial distribution of galaxies to hunt for the baryon acoustic oscillation signature. In about a year, the first phase of this analysis should be complete, and we can begin to look for clues that reveal the nature of the universe's expansion.

In the meantime, the experiment has made some interesting astrophysical findings, such as the discovery of 16 ultrafaint dwarf galaxy candidates in the Milky Way's backyard. These very nearby galaxies contain as little as a few tens of stars and are among the most dark matter–dominated objects known in the universe. Their darkness makes them very hard to detect, but they are of interest as the building blocks of larger galaxies like our own Milky Way and as potential sites for probing the nature of dark matter.

More DES data are coming in all the time. As you read this, scientists are analyzing these observations for clues to the nature of dark energy. We do not yet know whether the DES will provide definitive answers—dark energy or modified gravity? vacuum energy or quintessence?—but we do know it will take the next major step in the hunt for dark energy and for the root cause of our universe's mysteriously accelerating expansion.

--Originally published: Scientific American 24(4); 46-53 (Winter 2015).

# The Ultimate X-Ray Machine

by Nora Berrah and Philip H. Bucksbaum

An atom, molecule or bacterium placed at the focal point of the world's most powerful x-ray laser doesn't stand a chance. Up to a trillion high-energy photons, moving in unison, sweep through the matter, heating it to more than one million degrees Celsius—hot as the solar corona—in less than a trillionth of a second. When a pulse of such extreme radiation hits neon atoms, all 10 electrons boil off each atom, and the denuded nuclei explode away from their similarly ionized neighbors. The obliteration leaves a trail of destruction that can illuminate some of the mysteries of nature.

In the case of exploding neon, for instance, the x-ray laser astonishingly strips away the atoms' electrons from the inside out. The electrons orbit each nucleus in onionlike shells, but the outer shells are nearly transparent to x-rays. So the beam continues on until it hits the two electrons in the innermost shell. They take the brunt of the radiation, much as coffee in a microwave oven warms long before the cup that holds it. Sprung from the center, the two electrons leave behind a hollowed-out atom. But within femtoseconds (quadrillionths of a second), other electrons move inward to fill the gap. The cycle of hollowing and filling repeats—all within a single, exceedingly brief pulse of x-rays—until no electrons are left.

The result is an exotic, ionized plasma in a state called warm dense matter—it is usually found only in extreme settings such as nuclear fusion reactions and the cores of giant planets. For just a few femtoseconds, the destructive environment at the focus of an x-ray laser beam has no parallels on Earth.

The x-ray laser itself is as remarkable as the exotic phenomena it reveals. Known as the Linac Coherent Light Source (LCLS) at the SLAC National

Accelerator Laboratory, it evokes memories of the 1980s-era “Star Wars” missile-defense system. Advocates of that scheme proposed wielding x-ray lasers to shoot down ballistic missiles and satellites.

This real-world x-ray laser actually owes much more to the nation’s premier atom smashers built during that era—in particular, the SLAC linear accelerator, operated by Stanford University for the U.S. Department of Energy. That accelerator produced many of the discoveries and Nobel Prizes that kept the U.S. at the forefront of elementary particle physics for decades. And then it was rebuilt into the LCLS x-ray laser, which came online in October 2009.

Since then, the LCLS has become to atomic and plasma physics, chemistry, condensed matter physics and biology what the Large Hadron Collider at CERN near Geneva has been to elementary particle physics. The device gives physicists a formidable hammer to smash the building blocks of nature, creating from the debris new forms of matter, such as hollow atoms. We also use it like a powerful, high-speed microscope to zoom in on the quantum realm. The LCLS’s x-ray pulses are not only exceedingly bright but also incredibly short—just a few femtoseconds long. We use them to make atoms freeze in their tracks, to observe chemical reactions in progress, and to image living microbes and viruses in exquisite detail.

**Shadows of Atoms**

The x-ray laser fuses two of the main tools used by today’s experimental physicists: synchrotron light sources and ultrafast lasers. Synchrotrons are racetrack particle accelerators. Electrons circling through them throw off x-rays, which enter instruments arrayed around the circumference of the machines like pinwheel spokes.

One of us (Berrah) has spent a career using synchrotron x-rays to study the deep interior of atoms, molecules and nanosystems. X-ray light is ideal for this purpose. Its wavelengths are atomic size, so atoms cast a shadow in an x-ray beam. In addition, x-rays can be tuned to pick out specific kinds of atoms—say, only those of iron—and show where they sit in a solid or in a large molecule such as hemoglobin.

X-rays from synchrotrons are limited in one crucial way, however: they cannot trace out atomic motion inside most molecules or solids. The pulses

are not short enough or bright enough, so they produce only dim, blurry images unless the target is a crystal, where local forces hold millions of molecules in precise ranks like identical soldiers at attention.

Lasers, for their part, are far brighter because the light they emit is coherent. The electromagnetic field in a laser is not choppy, like the surface of a rough sea; it is a regular series of smooth oscillations.

Lasers can exploit that regularity to focus enormous amounts of energy onto a tiny spot of both space and time; they can switch on and off in as little as a femtosecond. One of us (Bucksbaum) uses ultrafast optical laser pulses as a strobe light to study the motion of atoms and the steps in chemical reactions.

But conventional lasers operate at or near visible wavelengths—wavelengths at least 1,000 times longer than those needed to resolve atoms. Just as weather radar can see a rainstorm but not resolve the raindrops, optical lasers can see how collections of atoms are moving, but they cannot distinguish individual atoms. To cast a sharp shadow, the wavelength of the light must be no bigger than the object under observation. So we need x-ray lasers to image atoms. Actually building such a device is no easy task, however.

**Death Rays**

At one time, the idea of building an x-ray laser seemed outlandish. Standard lasers are hard enough to construct. They work because atoms are like miniature batteries: they can absorb, store and release small amounts of energy in the form of photons. Typically atoms emit photons spontaneously, but early in the 20th century Albert Einstein discovered a way to trigger the release, a process known as stimulated emission. If you cause an atom to absorb a certain amount of energy and then hit it with a photon having the same amount of energy, the atom can release the originally absorbed energy—thus producing a clone of the photon. The two photons (the original one and its clone) go forth to trigger the release of energy from a pair of other atoms, and so on, building up a clone army in an exponential chain reaction. Laser beams are the result.

Even when conditions are right, atoms do not always clone photons as you might expect. The probability that an energetic atom will emit a photon

when hit by another is small compared with the likelihood that the atom will simply release its energy spontaneously. Conventional lasers overcome this limitation in two ways. They pump in energy to prime the atoms. And they use mirrors to send the cloned light surging back and forth, picking up new recruits along the way.

In a typical helium-neon laser used in supermarket price scanners, for example, a continuous stream of electrons collides with atoms in the gas, energizing it. And light bounces back and forth between mirrors 200 times before it exits the laser.

For an x-ray laser, every step of this process becomes much more diff cult. An x-ray photon may contain 1,000 times more energy than an optical photon does, so each atom must absorb 1,000 times more energy. The atoms do not hold on to that energy for long. Moreover, x-ray mirrors are hard to come by. Al though these impediments are not insurmountable, it takes an enormous input of energy to create the lasing conditions.

In fact, the first x-ray laser got its energy from an underground nuclear bomb test in the 1980s. It was built for a secret project, code-named Excalibur, carried out by Lawrence Livermore National Laboratory, east of San Francisco. The project is still classified, although some information about it has been made public. The device was a component of former president Ronald Reagan's Strategic Defense Initiative (aka "Stars Wars") and was meant to act as a death ray to shoot down missiles and satellites.

During the same decade, Lawrence Livermore also built the first nonnuclear, laboratory-scale version of an x-ray laser. It was powered by giant optical lasers originally designed to test properties of nuclear weapons. But these early devices were not practical research instruments. For decades the possibility that x-ray lasers would ever be used routinely for science applications seemed remote.

---

## Anatomy of the X-ray Laser

The LCLS is the closest thing to a starship laser blaster that earthlings have ever created. Its engine is a linear particle accelerator—a gigantic version of the electron guns inside old-style TV sets—that fires electrons at near light speed. The heart of the machine is the undulator, which causes the electrons to zigzag. That in turn creates synchrotron radiation—in this case, x-rays. Because the electrons are moving nearly as fast as the x-rays they produce, the process feeds on itself and produces an unusually pure and intense beam.

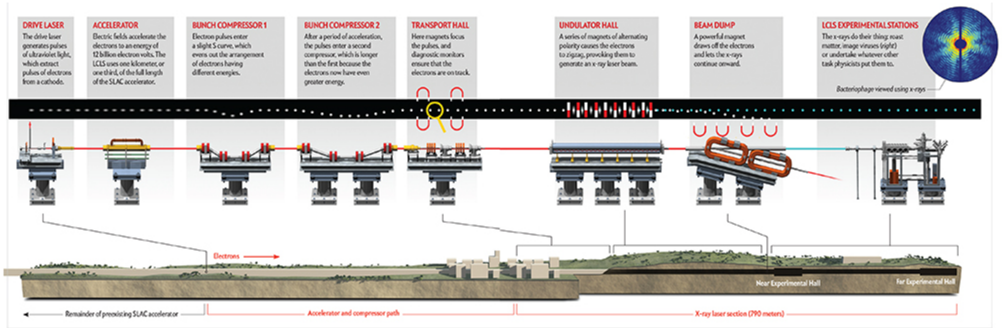



*Illustration by Don Foley*

---

## Subatomic Demolition Derby

The breakthrough that finally enabled investigators to develop x-ray lasers for civilian use came from another Bay Area institution, using a device intended for a different purpose entirely. In the 1960s Stanford built the world's longest electron accelerator. At three kilometers long, the building can be seen from space—it looks like a needle pointing from the mountains to the heart of the university's campus. The SLAC linac (short for linear accelerator) boosts dense bunches of electrons to velocities extremely close to the speed of light—so close that in a one-second race, the photons would travel nearly 300 million kilometers, and the electrons would trail them by just one centimeter.

The SLAC machine led to three Nobel Prizes for experimental discoveries in particle physics, but eventually it reached the end of its useful life. Particle physicists now make their discoveries at the Large Hadron Collider. A decade ago Stanford and SLAC's parent agency—the Department of Energy's Office of Science—decided to turn part of the aging machine into an x-ray laser by outfitting it with the same device used to produce x-rays at modern synchrotrons: an undulator.

Undulators consist of a series of magnets that generate alternating magnetic fields. Electrons moving through undulators wiggle and emit x-rays. In synchrotrons, which are closed loops, electrons bend into arced trajectories once they leave the undulator. That way the particles get out of the way, and the x-rays can move unimpeded to experimental stations. The electrons keep circling the racetrack, emitting a burst of x-rays each time they pass through the undulator.

The SLAC accelerator, however, is a straight line, and the un dulator is unusually long (130 meters). The electrons move along the same path as the photons do—and at nearly the same speed. The result is a subatomic demolition derby. The electrons cannot get out of the way of the x-ray photons they have emitted, so the photons sideswipe them again and again. In so doing, the photons induce the electrons to emit clone x-ray photons through the process of stimulated emission.

The process is analogous to what happens in an optical laser, but with a difference. Mirrors are not needed to bounce the light back and forth through the electrons, because they travel together. All it takes to produce the laser is an intense beam of fast electrons and a space big enough to house a long undulator. And SLAC possesses both.

If everything is lined up nearly perfectly, voilà, an extraordinarily bright x-ray beam. At the end of the line, the electrons are diverted, and the photons enter the experimental stations. The system is known technically as a free-electron laser.

Though not a gun for “Star Wars,” the LCLS is still a formidable device. Its peak focused intensity, 1018 watts per square centimeter, is billions of times greater than that of synchrotron light sources. The laser can cut through steel. Its oscillating electromagnetic field can be 1,000 times stronger than the fields that bind atoms to one another in molecules.

**Peering into Jupiter's Core**

Demand for the laser is so great that it can accommodate fewer than one in four research proposals to use it. On-site staff scientists work with large visiting teams of students, postdocs and senior scientists in intense marathons, 12 hours a day for five days. Every microsecond counts.

The research made possible by x-ray lasers is broad—and not limited by the conventional boundaries of physics. Just this year one collaboration that included biologists reported using the LCLS to make the first high-resolution x-ray images of living bacteria, and another pieced a series of x-ray snapshots together to create a stunning 3-D model of a virus.

In our own research in atomic, molecular and optical physics, we have used the LCLS recently to investigate two scientific problems that

particularly interest us. The first is how matter behaves under extreme conditions. The second is what can be learned from the ultrafast imaging of molecules.

Recall those strange, hollowed-out neon atoms we described earlier. It takes mere femtoseconds for the electrons from the outer shells of an atom to fall in to replace those that have been lost from the inner shells (a phenomenon called Auger relaxation). If we shine a shorter, one-femtosecond x-ray pulse on the atom, no outer electrons will have time to drop into the vacant inner shell. While the atoms are hollow, they should thus be transparent to any additional x-ray radiation, no matter how intense. And we have in fact detected this hollow transparency at the LCLS—not only in atoms but also in molecules and larger bits of matter.

Theorists suggest that inside giant planets such as Jupiter, temperatures reach 20,000 degrees Celsius—four times hotter than the surface of the sun. Hydrogen and helium, the planet's main constituents, presumably take on exotic solid phases having extreme densities and structures. Yet little is known about the specifics. Even the strength of the material, or its compression in response to pressure, is not easy to measure and not well understood from basic principles. So far research in this domain has relied heavily on theoretical models. Experiments that can validate the models have been scarce.

Some of the first experiments done at the LCLS attempted to re-create these hostile conditions. The laser's ultrahigh intensity can heat matter with dizzying speed, producing unusual effects. For instance, we observed for the first time how multiple x-rays can gang-tackle molecules made of many atoms to liberate electrons that are strongly bound to atomic nuclei, a process called multiphoton absorption.

Bright x-rays can, in addition, rapidly break all the bonds in molecules that are expected to reside inside giant planets, including water, methane and ammonia. Measurements of matter in extreme conditions induced by x-rays have helped determine the equation of state—the formula that governs the density, temperature and pressure—in cores of giant planets and during meteor impacts.

**Exploding Proteins**

The second line of research—exploiting the laser as an x-ray high-speed camera to image molecules and record movies of physical, chemical and biological dynamics—is filling in a serious gap in our knowledge. Researchers know distressingly little about the structure of many biological molecules—in particular, membrane proteins and large macromolecular complexes. The standard technique, crystallography, starts by growing a crystal that is large enough and perfect enough to diffract a beam of synchrotron x-rays. The resulting pattern reveals the structure of the molecule.

Unfortunately, x-rays rapidly damage the molecules they are probing. To compensate, researchers must prepare large crystals, yet many molecules of interest, including membrane proteins, are very difficult to crystallize. The synchrotron technique is also slow and thus unable to observe transient phenomena that occur on the femtosecond chemical timescale.

At first glance, the LCLS seems exactly the wrong tool for the job. Because it is billions of times more intense than synchrotron light sources, fragile materials such as proteins or noncrystalline systems cannot survive even one pulse of its x-rays before they explode and turn into a very hot soup of plasma.

Ironically, that destructive intensity is just what we need. Because the pulse is so short and bright, it can capture an image faster than the molecule is able to blow up. Consequently, although the laser obliterates the sample, it captures a clear image of the molecule just before its demise.

This concept, called diffraction before destruction, is already beginning to pay off. Scientists have used femtosecond crystallography to record diffraction patterns of nanocrystals, proteins and viruses. Recent work has mapped out the structure of proteins involved in sleeping sickness, a fatal disease caused by protozoan parasites.

Now that the LCLS has pioneered the technology, other free-electron x-ray lasers are in the works. In Japan, the SACLA laser facility opened in 2011. In Europe, a large x-ray laser is being built near Hamburg, Germany, and is scheduled for completion in 2017. In the U.S., SLAC is building the LCLS II. This upgrade will provide soft x-rays at a high repetition rate, which will let us conduct new kinds of experiments.

This new generation of machines has been designed to create a more stable, better controlled and more intense laser beam. One particularly important goal is to make the x-ray pulses even shorter. By using pulses as short as 0.1 femtosecond (100 attoseconds, or quintillionths of a second), we might begin to observe the motion not just of atoms but also of electrons within atoms and molecules. New devices could even allow us to control this motion. The dream of making movies that show how chemical bonds break and how new ones form is within our reach.

--Originally published: Scientific American 24(4); 54-61 (Winter 2015).

# SECTION 3

# Hidden Worlds

# Mystery of the Hidden Cosmos

by Bogdan A. Dobrescu and Don Lincoln

The beautiful spinning pinwheel of the Andromeda galaxy, our celestial neighbor, poses a mystery. The breakneck speed of its rotation cannot be explained by applying the known laws of physics to the disk's visible matter. By rights, the gravity generated by the galaxy's apparent mass should cause the stars in the periphery to move more slowly than they actually do. If there were nothing more to the galaxy than the visible matter, then Andromeda—and nearly all such fast-rotating galaxies—simply should not exist.

Cosmologists believe that some unseen kind of matter surrounds and permeates Andromeda and other galaxies, adding the necessary gravitational force to keep them spinning as observed. This dark matter appears to contribute about 25 percent of the universe's mass. And it could also explain other puzzling aspects of the cosmos, from the exceedingly fast motion of galaxies within clusters of galaxies, to the distribution of matter that arises when clusters collide, to the bending of light by the gravity of distant galaxies, a phenomenon known as gravitational lensing.

The simplest theories of dark matter postulate that a single kind of particle contributes the unseen mass. But in spite of decades of searching for direct evidence of the dark matter particle, no one has been able to prove its existence. A few discrepancies also remain between astronomical observations and this simple theory.

Some scientists have begun to question the traditional, single-particle theories and to imagine a more complicated form of dark matter. Perhaps, they suggest, a wide array of dark species exist. After all, ordinary matter comes in many forms—maybe dark matter is similarly complex.

Over the past few years both theoretical work and observations of colliding galaxies have lent preliminary support to the idea that dark matter comes in several varieties. Even more intriguing, these advances hint that previously unsuspected forces act strongly on dark matter and very feebly (or not at all) on ordinary matter. Such forces could help explain some of the discrepancies between the basic dark matter model and observations. If complex dark matter did exist, it would make for a more interesting and intricate universe than cosmologists usually imagine.

**Hidden Matter**

Although we do not yet know what constitutes dark matter, we do know something about its properties from our observations of how it influences normal matter and from simulations of its gravitational effects. We know, for instance, that dark matter must be moving much slower than the speed of light. Otherwise the density fluctuations present in the early universe would not lead to the galactic structures observed today. We also know that dark matter must be electrically neutral because it does not absorb or emit electromagnetic radiation.

The particles that compose dark matter are probably massive, or else they would have to be moving near the speed of light, which data from the early universe rule out. They cannot interact via the strong force, which binds atomic nuclei together; otherwise we would have seen evidence in dark matter's interaction with high-energy charged particles called cosmic rays.

Until recently, scientists believed that dark matter might interact via the weak force (responsible for radioactive decay), but new observations have undercut that notion. If dark matter did experience weak force interactions, it appears there would have to be other undiscovered particles of regular matter as well.

Dark matter must be stable on cosmic time scales, for the simple reason that there is no credible mechanism to continually produce dark matter. The stuff thus must have originated in the primordial big bang. That in turn implies something profound: the stability of dark matter over billions of years tells us that it possesses a property that is "conserved," meaning it cannot change. And it rules out the possibility that dark matter particles could decay because doing so would alter the conserved property.

The situation is analogous to the familiar electrical charge, which ensures that the electron is stable. It is a truism of physics that particles decay in to lighter ones unless something prevents that decay. The electron is electrically charged, and the only known stable particles lighter than it are electrically neutral: the photon and the neutrinos. Energy considerations would allow the electron to decay into these objects, but because conservation of charge prohibits such decays, the electron stays an electron.

In most dark matter theories, the conserved quantity of dark particles is called parity, for historical reasons. The dark matter particle has a parity of –1; all other known particles have a parity of +1. If a dark matter particle decayed into ordinary matter, the parity would not be conserved. So the theories assume that dark particles are prohibited from decaying.

The simplest theory that meets all the conditions physicists have outlined posits a single particle responsible for dark matter. They call it a WIMP, a weakly interacting massive particle. (The term “weakly” here is used in the generic sense and does not necessarily mean the weak nuclear force.) WIMPs make sense for many theoretical reasons, but they are proving harder to find than many physicists expected.

Since the 1990s scientists have been running various experiments aimed at directly detecting WIMPs through their very rare interactions with ordinary matter. To achieve the necessary sensitivity, the detectors are cooled to extremely low temperatures. They are also buried deep underground to shield them from ubiquitous cosmic rays, which can mimic a dark matter signature. Yet despite increasingly powerful experiments, no conclusive sign of WIMPs has emerged.

Although the WIMP model does explain many aspects of our observed universe, it does not account for everything. For example, WIMP theories predict that a much greater number of small satellite galaxies should be orbiting the Milky Way than do actually appear to swirl around it. The theories also predict that dark matter should be denser in the center of galaxies than it seems to be, based on the galaxies’ observed rotation rates. The situation is evolving rapidly, however. The recent discovery of additional satellite galaxies by the Dark Energy Survey collaboration suggests the problem with the Milky Way’s dwarf galaxies may simply be that many have yet to be found.

Nevertheless, these WIMP shortcomings have left the door open for more unconventional dark matter models.

**Complex Dark Matter**

It could be that there is more than just one kind of dark particle. An alternative possibility is that several classes of dark matter particles exist, as well as a variety of forces that act only on them. One idea that appears to reconcile all the observations and simulations is the possibility that dark matter particles interact with one another through some force that ordinary matter cannot feel.

These particles could, for instance, carry a new kind of “dark charge” that attracts or repels them even though they are electrically neutral. Ordinary particles with electrical charge can emit photons (particles of light that are the carriers of the electromagnetic force). Perhaps dark particles have dark charge and can emit “dark photons”—not particles of light but rather particles that interact with dark charge in the same way that photons interact with electrical charge.

The parallels to the world of normal matter must end at a certain point, however. If rules of the dark world exactly mirrored ours, then dark atoms would form and emit dark photons at the same rate that ordinary matter emits ordinary photons. We know from observing the shapes of galaxies that this does not happen.

Photon emission and galaxy shape might not seem connected, but they are. It is through the emission of photons that clouds of gas inside galaxies radiate electromagnetic energy. That radiation results in the spinning matter inside the clouds clumping together and eventually relaxing into a disklike structure.

If the rules and forces governing the behavior of dark matter were the same as ours, the emission of dark photons would result in all dark matter galaxies forming flattened disks. Yet we know that the distribution of most of the dark matter required to explain our familiar galaxies is more like a spherical cloud. We can thus rule out an exact mirror world of dark matter.

Still, many alternatives remain. It is possible, for example, that a small fraction of dark matter mirrors the rules of our universe and that the

remainder acts more like the simple WIMPs. Or perhaps the dark charge is effectively much smaller than the electrical charge of our electrons and protons, resulting in a far smaller amount of dark photon emission.

Theorists, including one of us (Dobrescu), are generating many ideas about possible particles and forces of the dark sector, using existing data to guide our thinking and constrain speculations. One of the simplest scenarios involves just two kinds of dark matter particles. It offers a glimpse of some of the physics that could operate in complex dark matter.

## Dark Photons

Imagine a dark world in which two kinds of dark charge exist—one positive and one negative. In this world, a form of dark electromagnetism allows the dark matter particles to both emit and absorb dark photons. The dark particles are charged in a way analogous to ordinary electrons and antielectrons (aka positrons). Positively and negatively charged dark matter particles should thus be able to meet and annihilate into dark photons, just as electrons and antielectrons annihilate on contact and convert their combined mass into an equivalent amount of photons.

We can make some conclusions about the strength of the dark electromagnetism force—and thus how often dark matter annihilation occurs—by considering how this force would affect galaxies. Recall that the reason galaxies have a flattened structure is that electromagnetism allows ordinary matter to lose energy and settle into disks. This energy loss occurs even without annihilation. Because we know that dark matter is primarily distributed spherically around most galaxies and does not collapse to a disk, we can conclude that it cannot lose energy via dark photon emission at the same rate that ordinary matter does.

In a study published in 2009, Lotty Ackerman, Matthew R. Buckley, Sean M. Carroll and Marc Kamionkowski, all then at the California Institute of Technology, showed that this requirement implies that the dark charge must be very small, about 1 percent of the value of the electrical charge. Yet even at such a low value, the force could still exist and have significant effects on the structure of galaxies.

## Dark Galaxy

So far we have described a version of dark matter consisting of a charged dark particle and its oppositely charged antiparticle emitting dark photons. But this scenario still pales in comparison to the complexity of ordinary matter. What would a dark matter world having multiple different charged particles look like?

Many theories of complex dark matter include two or more hypothetical dark particles. One particularly intriguing example was proposed in 2013 by JiJi Fan, Andrey Katz, Lisa Randall and Matthew Reece, all then at Harvard University, who referred to their model as "partially interacting dark matter." They assumed the bulk of dark matter is made up of WIMPs. But they also postulated that a small component consists of two classes of dark fermions. (Fermions are particles—such as protons, neutrons and the quarks that compose them—that have a quantum-mechanical spin of ½.) One dark fermion in this theory is heavy and the other is light, but both carry dark charge. They both thus emit dark photons and can be attracted to each other.

The proposed situation is broadly similar to postulating a dark proton, a dark electron and a dark photon to carry the dark electromagnetism that binds them together, although one must take care to not overinterpret the correspondence. If the dark fermions had appropriate masses and charges, they could conceivably combine to create dark atoms with their own dark chemistry, dark molecules and possibly even more complex structures. This concept of dark atoms had already been explored in detail in 2010 by David E. Kaplan, Gordan Z. Krnjaic, Keith R. Rehermann and Christopher M. Wells, all then at Johns Hopkins University.

The Harvard physicists who proposed dark matter fermions went on to derive an upper limit on the fraction of dark matter that may be strongly interacting with dark photons, given the constraints imposed by astronomical observations. They determined that the total mass of such particles might equal that of all visible matter.

In this model, the Milky Way galaxy consists of a large spherical cloud of WIMP-like dark particles, which contributes 70 percent of the total matter, encircling two flattened disks, each containing 15 percent of the matter. One disk is normal matter, which includes the spiral arms that we can see, and the other consists of strongly interacting dark matter. The two disks need not be exactly aligned, but they would have a similar orientation.

In this picture, a dark matter galaxy basically coexists in the same space as our familiar Milky Way. A cautionary note: the dark matter galaxy could not include dark stars or large planets. If it did, they would have caused gravitational lensing effects on the light from ordinary matter; such effects have not been seen.

The idea may sound radical, but the extra disk in our galaxy would do little to change the normal matter cosmos with which it coexists. After all, to be correct, any theories about dark matter have to be consistent with existing observations of visible matter. We could be living in such a universe without even knowing it.

**Experimental Prospects**

Scientists can search for complex dark matter in the same ways they search for WIMPs: by using sensitive underground detectors. One consequence of the Harvard model of partially interacting dark matter is that any of these dark fermions passing through our detectors would be denser than that predicted in WIMP models. If correct, the increased density could mean that the probability of finding dark matter with these detectors is higher than conventional theory predicts.

The search for such dark particles is under way. Physicists are also conducting experiments with particle accelerators in the hope of making dark matter, along with all the other exotic particles generated by high-energy collisions. Because we know very little about how dark matter interacts with ordinary matter—and thus which particular processes inside the accelerator might give rise to it—the program of investigation is deliberately broad. It is sensitive to models ranging from the simple WIMP to a more complex dark sector.

Of course, we must make some assumptions. If dark matter interacts only gravitationally, for example, we will never create it in any conceivable accelerator, nor will we see it in any direct search—gravity is simply too weak. So scientists assume that dark matter interacts with ordinary matter via a force or forces that are much stronger than gravity and yet weak enough to not have been observed already. This force linking dark and ordinary matter, we should note, is thought to be different from the chargelike force through which dark matter might interact with itself.

The Large Hadron Collider (LHC) at CERN near Geneva is the world's highest-energy accelerator. That gives it an edge when searching for heavier versions of dark matter, for two reasons. First, the more massive a particle is, the more energy it takes to produce it in an accelerator. Second, dark matter particles may interact more frequently as their energy rises.

We already know that dark matter can interact only very weakly with ordinary matter. So we cannot expect to observe it directly in the detector, which is made of ordinary matter. Instead scientists have been searching for dark matter by looking for collisions in which energy is missing.

For example, two protons might collide and produce some ordinary particle or particles exiting one side of the collision and a couple of dark matter particles on the other. A detector would register energy on one side but nothing on the other side. Scientists calculate how many collisions would be expected to show this striking configuration if dark matter did not exist. They then check to see whether there are more than expected.

So far no signs of such an excess have shown up inside the LHC—an indication that dark matter's interactions with ordinary matter must be very infrequent, if they occur at all. But a new opportunity for seeing signs of dark matter began in June with the start of the LHC's upgraded, higher-energy second run. That means that the discovery of the century could be right around the corner.

The searches for dark matter we have just described are suitable for finding both WIMPs and complex dark matter. But other approaches in development aim more specifically at the complex dark sector. Many of them search for the dark photon.

Some models suggest that dark photons can continually transform into ordinary photons and back again via the laws of quantum mechanics, potentially presenting an opportunity to see the photons that result. Other models suggest that certain dark "photons" have a nonzero mass (quite different from the familiar massless photon). If a dark photon has mass, it can potentially decay into lighter particles. And if a dark photon can indeed transform briefly into a normal photon, then there is a small chance that during the transformation process it can produce pairs of electrons and

antielectrons (also known as positrons). Alternatively, it might create a muon (a cousin of the electron) and an antimuon.

Experimental collaborations, including a project on which one of us (Lincoln) is a member, are now monitoring accelerator collisions for the production of electron-antielectron pairs or muon-antimuon pairs. In addition to studies at the LHC, such work is ongoing as part of the KLOE-2 project at the Frascati National Laboratories at the National Institute for Nuclear Physics in Italy, the Heavy Photon Search experiment at the Thomas Jefferson National Accelerator Facility in Newport News, Va., and the BaBar detector experiment at SLAC National Accelerator Laboratory in Menlo Park, Calif. Scientists are even digging through data taken more than a decade ago by a SLAC experiment known as mQ.

Our group at Fermi National Accelerator Laboratory in Batavia, Ill., is trying to make beams of dark matter particles by shooting intense beams of neutrinos at distant detectors. Neutrinos are very light subatomic particles that interact essentially exclusively through the weak nuclear force. If dark matter interacts with ordinary matter via particles like dark photons, it is possible that dark matter is being made in the same beams and can possibly be detected in Fermilab's MiniBooNE, MINOS or NOvA detectors.

Finally, scientists can search for astronomical signs that dark matter is interacting in situations such as when galaxies collide. In such scenarios, when the dark matter from one galaxy slams into the dark matter in another, the particles could repel one another by exchanging dark photons. Several studies of galaxy crashes have failed to find evidence of this phenomenon. But recent observations of the cluster Abell 3827, which is particularly close to Earth and well oriented, hint at just such a pattern. Further observations of that and other galaxy collisions will be necessary to confirm the signal, but the data from this cluster so far look promising for complex dark matter models.

## A Cosmic Stumper

There is no question that we are facing a profound conundrum. On large scales, ordinary gravitationally bound matter does not act in ways consistent with the known laws of physics and the observed distribution of mass. Because of this disagreement, most scientists are confident that some form

of dark matter exists. What form this matter takes, however, has become increasingly contentious as our experiments repeatedly fail to find evidence for the simplest dark matter models. For this reason—and because of some persistent discrepancies between the simple WIMP model predictions and astronomical observations—theories of complex dark matter are becoming more appealing. These models offer theorists more parameters to tune and thus to improve the agreement between data and theory. They also more closely match the variation and richness of normal matter.

A criticism of this approach may be that it works overly hard to keep the dark matter hypothesis alive. Could this situation be similar to the discredited idea of epicycles, whereby 16th-century astronomers tried to save Earth's central position in the universe by continually tweaking a fatally flawed theory? We think not. Dark matter explains many astronomical conundrums remarkably well, and there is no a priori reason why dark matter should be as simple as the WIMP hypothesis.

The real message is that we have a mystery before us. We do not know what the answer will be. Until we find it, we must be open to myriad explanations, including the fascinating possibility that we might be living alongside a dark parallel reality. Could it be that a dark matter scientist has turned its attention to its skies and is wondering about us?

---

## All the Matter in the Universe

In addition to the normal "baryonic" matter in the cosmos, some hidden form of matter must be out there, gravitationally tugging on galaxies to keep them spinning as fast as they do and holding clusters of galaxies together. Yet no direct evidence exists to explain what dark matter is. The likely culprit is one or more species of particle that do not feel the electromagnetic or strong forces and thus do not emit or reflect light or bind to atomic nuclei. Exactly what form these particles take, however, is an open question.

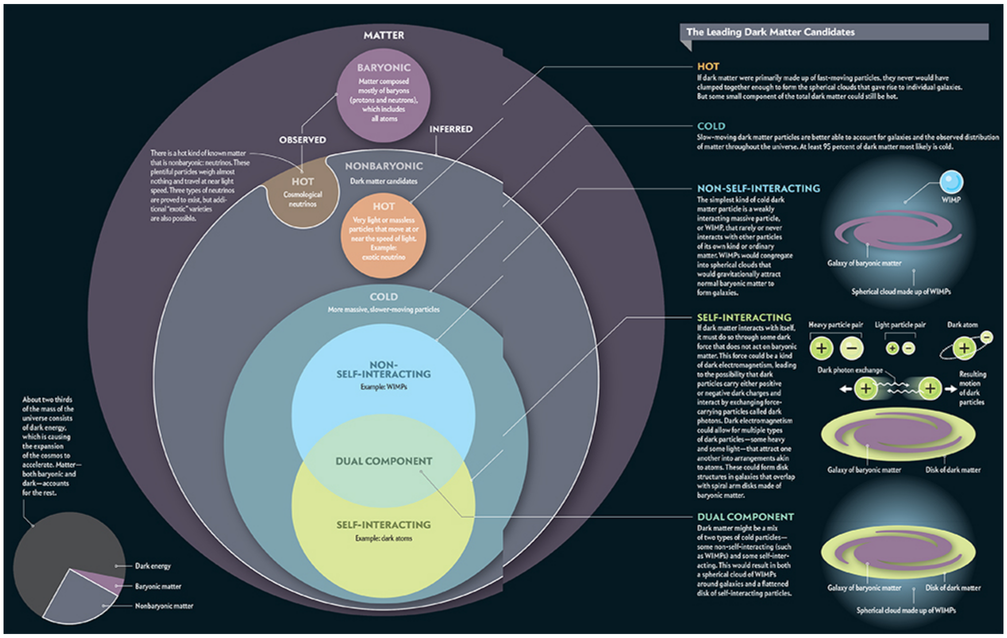



*Illustration by Jen Christiansen*

---

--Originally published: Scientific American 24(4); 62-69 (Winter 2015).

# Does the Multiverse Really Exist?

by George F. R. Ellis

In the past decade an extraordinary claim has captivated cosmologists: that the expanding universe we see around us is not the only one; that billions of other universes are out there, too. There is not one universe—there is a multiverse. In *Scientific American* articles and in books such as Brian Greene's *The Hidden Reality*, leading scientists have spoken of a super-Copernican revolution. In this view, not only is our planet one among many, but even our entire universe is insignificant on the cosmic scale of things. It is just one of countless universes, each doing its own thing.

The word "multiverse" has different meanings. Astronomers are able to see out to a distance of about 42 billion light-years, our cosmic visual horizon. We have no reason to suspect the universe stops there. Beyond it could be many—even infinitely many—domains much like the one we see. Each has a different initial distribution of matter, but the same laws of physics operate in all. Nearly all cosmologists today (including me) accept this type of multiverse, which Max Tegmark calls "level 1."

Yet some go further. They suggest completely different kinds of universes, with different physics, different histories, maybe different numbers of spatial dimensions. Most will be sterile, although some will be teeming with life. A chief proponent of this "level 2" multiverse is Alexander Vilenkin, who envisions an infinite set of universes containing an infinite number of galaxies—and infinitely many people with your name who are reading this article.

Similar claims have been made since antiquity by many cultures. What is new is the assertion that the multiverse is a scientific theory, with all that implies about being mathematically rigorous and experimentally testable. I am skeptical about this claim. I do not believe the existence of those other

universes has been proved—or ever could be. Proponents of the multiverse, as well as greatly enlarging our conception of physical reality, are implicitly redefining what is meant by “science.”

## Over the Horizon

How would universes proliferate in a multiverse, and where would they all reside? Advocates have suggested several alternatives. The universes might be sitting in regions of space far beyond our own, as envisaged by the chaotic inflation model of Alan H. Guth, Andrei Linde and others. They might exist at different epochs of time, as proposed in the cyclic universe model of Paul J. Steinhardt and Neil Turok. They might exist in the same space we do but in a different branch of the quantum wave function, as advocated by David Deutsch. They might not have a location, being completely disconnected from our spacetime, as suggested by Tegmark and Dennis Sciama.

Of these options, the most widely accepted is that of chaotic inflation, and I will concentrate on it; however, most of my remarks apply to all the other proposals as well. The idea behind chaotic inflation is that space at large is an eternally expanding void, within which quantum effects continually spawn new universes like a child blowing bubbles. The concept of inflation goes back to the 1980s, and physicists have elaborated on it based on their most comprehensive theory of nature: string theory.

String theory allows bubbles to differ from one another in fundamental ways. In effect, each begins life not only with a random distribution of matter but also with random kinds of matter. Our universe contains particles such as electrons and quarks interacting through forces such as electromagnetism; other universes may have distinct particles and forces—thus, different local laws of physics. The full set of allowed local laws is known as the landscape. In some interpretations of string theory, the landscape is immense, ensuring a tremendous diversity of universes.

Many physicists who talk about the multiverse, especially advocates of the string landscape, do not care much about parallel universes per se. For them, objections to the multiverse as a concept are unimportant. Their theories live or die based on internal consistency and, one hopes, eventual laboratory testing. They assume a multiverse context for their theories

without worrying about how it comes to be—which is what concerns cosmologists.

For a cosmologist, the basic problem with all multiverse proposals is the presence of a cosmic visual horizon. The horizon is the limit to how far away we can see, because signals traveling toward us at the speed of light (which is finite) have not had time since the beginning of the universe to reach us from farther out. All the parallel universes lie outside our horizon and remain beyond our capacity to see, now or ever, no matter how technology evolves. In fact, they are too far away to have had any influence on our universe whatsoever. That is why none of the claims made by multiverse enthusiasts can be directly substantiated.

Proponents tell us that one can state in broad terms what happens 1,000 times as far as our cosmic horizon, $10^{100}$ times, $10^{1,000,000}$ times, an infinity—all from data we obtain within the horizon. It is an extraordinary extrapolation. Maybe the universe closes up on a very large scale, and there is no infinity out there. Maybe all the matter in the universe ends somewhere, with empty space forever after. Maybe space and time come to an end at a singularity that bounds the universe. We just do not know, because we have no information about these regions. And we never will.

---

## What Lies Beyond?

When astronomers peer into the universe, they see out to a distance of about 42 billion lightyears, our cosmic horizon, which represents how far light has been able to travel since the big bang (as well as how much the universe has expanded in size since then). Assuming that space does not just stop there and may well be infinitely big, cosmologists make educated guesses as to what the rest of it looks like.

**Level 1 Multiverse: Plausible** The most straightforward assumption is that our volume of space is a representative sample of the whole. Distant alien beings see different volumes, but all of these look basically alike, apart from random variations in the distribution of matter. Together these regions, seen and unseen, form a basic type of multiverse.

**Level 2 Multiverse: Questionable** Many cosmologists go further and speculate that, sufficiently far away, things look quite different from what we see. Our environs may be one of many bubbles floating in an otherwise empty background. The laws of physics would diff er from bubble to bubble, leading to an almost inconceivable variety of outcomes. Those other bubbles may be impossible to observe even in principle. The author and other skeptics feel dubious about this type of multiverse.

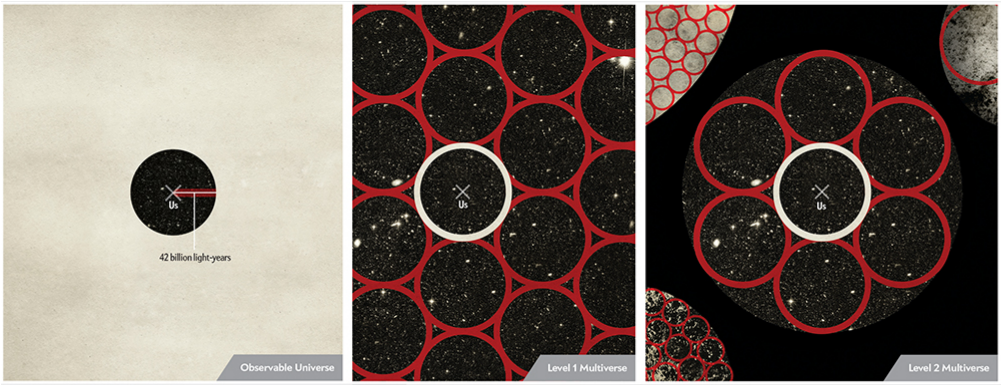



*Illustration by Chad Hagen*

### Seven Questionable Arguments

Most multiverse proponents are careful scientists who are quite aware of this problem but think we can still make educated guesses about what is going on out there. Their arguments hew to seven broad themes, each of which runs into trouble.

**Space has no end**. Few dispute that space extends beyond our cosmic horizon and that many other domains lie beyond what we see. If this limited type of multiverse exists, then we can extrapolate what we see to domains beyond the horizon, with our uncertainty increasing along with distance. More elaborate variations, including alternative physics out where we cannot see, are easy to imagine. The trouble with this type of extrapolation from the known to the unknown is that no one can prove you wrong.

How can scientists decide whether their picture of an unobservable region is a reasonable or an unreasonable extrapolation of what we see? If other universes have different initial distributions of matter, might they also have different values of fundamental physical constants, such as those that set the strength of nuclear forces? Different assumptions imply either or both.

**Known physics predicts other domains**. Proposed unified theories predict entities such as scalar fields, a hypothesized relative of other space-filling fields such as the magnetic field. Such fields should drive cosmic inflation and create universes ad infinitum. These theories are well grounded theoretically, but the nature of the hypothesized fields is unknown, and experimentalists have yet to demonstrate their existence, let alone measure their supposed properties. Crucially, physicists have not

substantiated that the dynamics of these fields would cause different laws of physics to operate in different bubble universes.

**The theory that predicts an infinity of universes passes a key observational test**. The cosmic microwave background radiation reveals what the universe looked like at the end of its hot, early expansion era. Patterns in it suggest that our universe really did undergo a period of inflation. But not all varieties of inflation continue on to create an infinite number of bubble universes. Observations do not single out the required type of inflation from other types. Some cosmologists such as Steinhardt even argue that eternal inflation would have led to different patterns in the background radiation than we see. Linde and others disagree. Who is right? It all depends on what you assume about the physics of the inflationary field.

**Fundamental constants are finely tuned for life**. A remarkable fact about our universe is that physical constants have just the right values needed to allow for complex structures, including living things. Steven Weinberg, Martin Rees, Leonard Susskind and others contend that an exotic multiverse provides a tidy explanation for this apparent coincidence: if all possible values occur in a large enough collection of universes, then ones that permit life are inevitable. This reasoning has also been used to explain the density of the dark energy that is speeding up the expansion of the universe today.

I agree that the multiverse is one possible explanation for the value of this density; it may even be the best option we have right now. But we have no hope of testing it observationally. And most analyses of the issue assume the basic equations of physics are the same everywhere—only the constants differ. But if one takes the multiverse seriously, this need not be so.

**Fundamental constants match multiverse predictions**. This argument refines the previous one by suggesting that the universe is no more finely tuned for life than it strictly needs to be. Proponents have assessed the probabilities of various values of the dark energy density. The higher the value is, the more probable it is, but the more hostile the universe would be to life. The value we observe should be just on the borderline of uninhabitability, and it does appear to be so.

We cannot apply a probability argument, however, if there is no multiverse; statistics simply are not applicable if only one universe physically exists. This argument thus assumes the desired outcome before it starts. Probability can probe the consistency of the multiverse proposal but cannot prove its existence.

**String theory predicts a diversity of universes**. In current string theory, almost anything is possible. It predicts that many essential properties of our universe are pure happenstance. If the universe is unique, those properties seem inexplicable.

How can we understand, for example, the fact that physics has precisely those highly constrained properties that allow life to exist? If the universe is one of many, those properties make perfect sense. Nothing singled them out; they are simply the ones that arose in our region of space. Had we lived elsewhere, we would have observed different properties, if we could indeed exist there. (Life would be impossible in most places. )

Unfortunately, string theory is not tried and tested; it is not even a complete theory. If we had proof that string theory is correct, its theoretical predictions could be a legitimate, experimentally based argument for a multiverse. We do not have such proof.

**All that can happen, does happen**. Some physicists and philosophers speculate that nature never chose to obey some laws and not others; all conceivable laws do apply somewhere.

The idea is in spired in part by quantum mechanics, which, as Murray Gell-Mann memorably put it, holds that everything not forbidden is compulsory. A particle takes all the paths it can; we see the weighted average of all those possibilities. Perhaps the same is true of the entire universe, implying a multiverse.

But astronomers have no chance of observing this multiplicity of possibilities. We cannot even know what the possibilities are. This proposal requires some unverifiable organizing principle that decides what is allowed and what is not—for example, that all possible mathematical structures must be realized in some physical domain (as Tegmark has proposed). Yet we have no idea what kinds of existence this principle entails, apart from the fact that it must include the world we see around us. And we have no

way whatsoever to verify the existence or nature of any such organizing principle. It is in some ways an attractive proposition, but its proposed application to reality is pure speculation.

---

Does the Glove Fit?

As evidence for a multiverse, proponents often cite the density of the dark energy that dominates our universe. The process of eternal inflation endows each universe in a multiverse with a random density of dark energy. Relatively few universes have zero or a low value; most have higher values (blue area). But too much dark energy tears apart the complex structures needed to sustain life (red area). So most habitable universes should have a middling density of dark energy (peak of overlap region)—and, lo and behold, our universe does. Multiverse skeptics, though, say this reasoning is circular: it holds only if you assume the multiverse to begin with. It is a consistency test, not a proof.

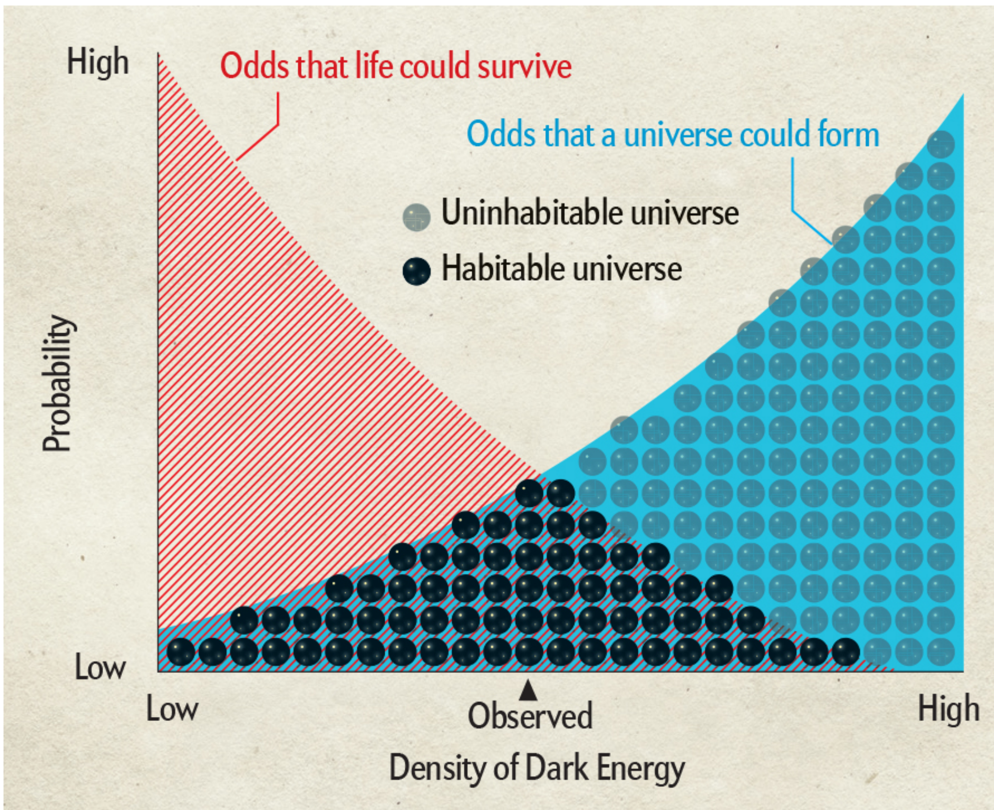



*Illustration by Jen Christiansen*

---

## Absence of Evidence

Cosmologists have also suggested various empirical tests for parallel universes. The cosmic microwave background radiation might bear some traces of other bubble universes if, for example, our universe has ever collided with another bubble of the kind implied by the chaotic inflation scenario. The background radiation might also contain signs of universes

that preceded the big bang in an endless cycle. These are indeed ways one might get real evidence of other universes. Some cosmologists have even claimed to see such signs. The observational claims are strongly disputed, however, and many of the hypothetically possible multiverses would not lead to such evidence. So observers can test only some specific classes of multiverse models in this way.

A second observational test is to look for evidence that one or more fundamental constants actually vary, which would corroborate the premise that the laws of physics are not so immutable after all. Some astronomers claim to have found such variations. But most consider the evidence dubious.

A third test is to measure the shape of the observable universe: Is it spherical (positively curved), hyperbolic (negatively curved) or “flat” (uncurved)? Multiverse scenarios generally predict that the universe is not spherical, because a sphere closes up on itself, allowing for only a finite volume. Unfortunately, this test is not a clean one. The universe beyond our horizon could have a different shape from that in the observed part. Moreover, not all multiverse theories rule out a spherical geometry.

The topology of the universe offers a better test: Does it wrap around, like a doughnut or pretzel? A wrapped universe is finite, so definitely inconsistent with the chaotic inflation scenario. A closed shape would produce recurring patterns in the sky, such as giant circles in the background radiation. Observers have failed to find any such patterns. But this null result just rules out specific types of a single universe—it does not require a multiverse.

Finally, physicists might hope to prove or disprove some of the theories that predict a multiverse. They might find observational evidence against chaotic versions of inflation. Recent observations by the Planck space observatory of directional unevenness in the cosmic microwave background radiation, for example, tend to call these models into question. Or they might discover a mathematical or empirical inconsistency that forces them to abandon the landscape of string theory. Either scenario would undermine much of the motivation for supporting the multiverse idea, al though it would not rule the concept out altogether.

## Too Much Wiggle Room

All in all, the case for the multiverse is inconclusive. The fundamental issue is the extreme flexibility of the proposal: it is more a concept than a well-defined theory. Most proposals involve a patchwork of different ideas rather than a coherent whole. The basic mechanism for eternal inflation does not itself cause physics to be different in each domain in a multiverse; for that, it needs to be coupled to another speculative theory. Although they can be fitted together, there is nothing inevitable about it.

The key step in justifying a multiverse is extrapolation from the known to the unknown, from the testable to the untestable. You get different answers depending on what you choose to extrapolate. Because theories involving a multiverse can explain almost anything, some multiverse variant can accommodate any observation. The various “proofs,” in effect, ask us to accept a theoretical explanation instead of insisting on observational testing. But such testing has, up until now, been the central requirement of the scientific endeavor, and we abandon it at our peril. If we weaken the requirement of solid data, we weaken the core reason for the success of science over the past centuries.

Now, a unifying explanation of some range of phenomena does carry greater weight than a hodgepodge of arguments for the same phenomena. If the unifying explanation necessarily assumes the existence of unobservable entities such as parallel universes, we might feel compelled to accept those entities.

But a key issue here is how many unverifiable entities are needed. Specifically, are we hypothesizing more or fewer entities than the number of phenomena to be explained? In the case of the multiverse, we are supposing the existence of a huge number—perhaps even an infinity—of unobservable entities to explain just one existing universe. It hardly fits 14th-century English philosopher William of Ockham’s stricture that “entities must not be multiplied beyond necessity.”

Proponents of the multiverse make one final argument: that there are no good alternatives. As distasteful as scientists might find the proliferation of parallel worlds, if it is the best explanation, we would be driven to accept it. Conversely, if we are to give up the multiverse, we need a viable

alternative. This exploration of alternatives depends on what kind of explanation we are prepared to accept. Physicists' hope has always been that the laws of nature are inevitable—that things are the way they are because there is no other way they might have been—but we have been unable to show this to be true. Other options exist, too. The universe might be pure happenstance—it just turned out that way. Or things might in some sense be meant to be the way they are—purpose or intent somehow underlies existence. Science cannot determine which is the case, because these are metaphysical issues.

Scientists proposed the multiverse as a way of resolving deep issues about the nature of existence, but the proposal leaves the ultimate issues unresolved. All the same issues that arise in relation to the universe arise again in relation to the multiverse. If the multiverse exists, did it come into existence through necessity, chance or purpose? That is a metaphysical question that no physical theory can answer for either the universe or the multiverse.

To make progress, we need to keep to the idea that empirical testing is the core of science. We need some kind of causal contact with whatever entities we propose; otherwise, there are no limits. The link can be a bit indirect. If an entity is unobservable but absolutely essential for properties of other entities that are indeed verified, it can be taken as verified. But then the onus of proof devolves to showing that it is absolutely essential to the web of explanation. The challenge I pose to multiverse proponents is: Can you prove that unseeable parallel universes are vital to explain the world we do see? And is the link essential and inescapable?

As skeptical as I am, I think that the contemplation of the multiverse is an excellent opportunity to reflect on the nature of science and on the ultimate nature of existence: why we are here. It leads to new and interesting insights and so is a productive research program. In looking at this concept, we need an open mind—though not too open. It is a delicate path to tread.

Parallel universes may or may not exist; the case is unproved. We are going to have to live with that uncertainty. Nothing is wrong with scientifically based philosophical speculation, which is what multiverse proposals are. But we should name it for what it is.

--Originally published: Scientific American 24(4); 70-75 (Winter 2015).

# Quantum Gravity in Flatland

by Steven Carlip

From its earliest days as a science, physics has searched for unity in nature. Isaac Newton showed that the same force responsible for the fall of an apple also holds the planets in their orbits. James Clerk Maxwell combined electricity, magnetism and light into a single theory of electromagnetism; a century later physicists added the weak nuclear force to form a unified "electroweak" theory. Albert Einstein joined space and time themselves into a single spacetime continuum.

Today the biggest missing link in this quest is the unification of gravity and quantum mechanics. Einstein's theory of gravity, his general theory of relativity, describes the birth of the universe, the orbits of planets and the fall of Newton's apple. Quantum mechanics describes atoms and molecules, electrons and quarks, the fundamental subatomic forces, and much besides. Yet in the places where both theories should apply—where both gravity and quantum effects are strong, such as black holes—they also seem incompatible.

Physicists' best efforts to create a single, unified theory that explains both quantum phenomena and gravity have failed miserably, giving answers that make no sense or no answers at all. Despite 80 years of work by physicists, including a dozen or so Nobel laureates, a quantum theory of gravity remains elusive.

Ask a physicist too hard a question, and a common reply will be, "Ask me something easier." Physics moves forward by looking at simple models that capture pieces of a complex reality. Researchers have worked on numerous such models for quantum gravity, including approximations that apply when gravity is weak or in special cases such as black holes.

Perhaps the most unusual approach is to neglect a whole dimension of space and work out how gravity would operate if our universe were only two-dimensional. (Technically, physicists refer to this situation as “(2+1)-dimensional,” meaning two dimensions of space plus one of time.) The principles that govern gravity in this simplified universe might also apply to our 3-D one, thus giving us some much needed clues to unification.

The idea of dropping down a dimension has a distinguished history. Edwin Abbott’s 1884 novel *Flatland: A Romance of Many Dimensions* follows the adventures of “A Square,” a resident of a 2-D world of triangles, squares and other geometric figures. Although Abbott intended it as a satirical commentary on Victorian society—*Flatland* had a rigid class hierarchy, with linear women at the bottom and a class of circular priests at the top—*Flatland* also triggered a surge of interest in geometry in diverse dimensions and remains popular today among mathematicians and physicists.

Researchers trying to wrap their minds around a higher-dimensional realm start by imagining what our 3-D world would look like to A Square. *Flatland* has also inspired physicists studying materials such as graphene that really do behave like 2-D spaces.

The first studies of Flatland gravity, made in the early 1960s, were a letdown. A 2-D space literally would not have enough room for changes in the gravitational field to propagate. In the late 1980s, however, the subject had a renaissance as researchers realized that gravity works in unexpected ways. It would still sculpt the overall shape of space and even create black holes.

Flatland gravity has been a case study in lateral thinking, letting us subject some of our speculative ideas, such as the so-called holographic principle and the emergence of time from timelessness, to a rigorous mathematical test.

**Time Management**

When physicists seek to develop a quantum theory of a force, we take the corresponding classical theory as our starting point and build on it. For gravity, that means general relativity, and there the trouble starts. General relativity involves a complex system of 10 equations, each with up to

thousands of terms. We cannot solve these equations in their full generality, so we face a daunting task in formulating their quantum version. But the mystery of why quantum gravity is so elusive is deeper still.

According to general relativity, the thing we call "gravity" is actually a manifestation of the shape of space and time. Earth orbits the sun not because some force tugs on it but because it is moving along the straightest possible path in a spacetime that has been warped by the sun's mass. Uniting quantum mechanics and gravity means somehow quantizing the structure of space and time itself.

That may not sound so challenging. Yet a cornerstone of quantum mechanics is the Heisenberg uncertainty principle, the idea that physical quantities are inherently fuzzy—fluctuating randomly and having no definite values unless they are observed or undergo an equivalent process. In a quantum theory of gravity, space and time themselves fluctuate, shaking the scaffolding on which the rest of physics is built. Without a fixed spacetime as the background, we do not know how to describe positions, rates of change or any of the other basic quantities of physics. Simply put, we do not know what a quantum spacetime means.

These general obstacles to conceptualizing quantized spacetime show up in several specific ways. One is the notorious "problem of time." Time is fundamental to our observed reality. Almost every theory of physics is ultimately a description of the way some piece of the universe changes in time. So we physicists had better know what "time" means, and the embarrassing truth is that we do not.

To Newton, time was absolute—standing outside nature, affecting matter but unaffected by it. The usual formulations of quantum mechanics accept this idea of an absolute time. Relativity, however, dethroned absolute time. Different observers in relative motion disagree about the passage of time and even about whether two events are simultaneous. A clock—as well as anything else that varies in time—runs more slowly in a strong gravitational field. No longer merely an external parameter, time is now an active participant in the universe. But if there is no ideal clock sitting outside the universe and determining the pace of change, the passage of time must arise from the internal structure of the universe. But how? It is hard to even know where to start.

The problem of time has a less famous cousin, the problem of observables. Physics is an empirical science; a theory must make verifiable predictions for observable quantities. In ordinary physics, these quantities are ascribed to specific locations: the strength of the electric field “here” or the probability of finding an electron “there.” We label “here” or “there” with the coordinates $x$, $y$ and $z$, and our theories predict how observables depend on the values of these coordinates.

Yet according to Einstein, spatial coordinates are arbitrary, human-made labels, and in the end the universe does not care about them. If you cannot identify a point in spacetime objectively, then you cannot claim to know what is going on at it. Charles Torre of Utah State University has shown that a quantum theory of gravity can have no purely local observables—that is, observables whose values depend on only a single point in spacetime. So scientists are left with nonlocal observables, quantities whose values depend on many points at once. In general, we do not even know how to define such objects, much less use them to describe the world we observe.

A third problem is how the universe came into being. Did it pop into existence from nothing? Did it split off a parent universe? Or did it do something else entirely? Each possibility poses some difficulty for a quantum theory of gravity. A related problem is a perennial favorite of science-fiction writers: wormholes, which form shortcuts between locations in space or even in time. Physicists have thought seriously about this idea—in the past 20 years they have written more than 1,000 journal articles on wormholes—without settling the question of whether such structures are possible.

A final set of questions revolves around the most mysterious beasts known to science: black holes. They may offer our best window into the ultimate nature of space and time. In the early 1970s Stephen Hawking showed that black holes should glow like a hot coal—emitting radiation with a so-called blackbody spectrum. In every other physical system, temperature reflects the underlying behavior of microscopic constituents. When we say a room is hot, what we really mean is that the molecules of air inside it are moving energetically. For a black hole the “molecules” must be quantum-gravitational. They are not literally molecules but some unknown microscopic substructure—what a physicist would call “degrees of

freedom”—that must be capable of changing. No one knows what they truly are.

### An Unattractive Model

At first glance, Flatland seems an unpromising place to seek answers to these questions. Abbott’s Flatland had many laws, but a law of gravity was not among them. In 1963 Polish physicist Andrzej Staruszkiewicz worked out what that law might be by applying general relativity. He found that a massive object in Flatland would bend the 2-D space around it into a cone, like a party hat made by twisting a flat piece of paper. A small object passing the apex of this cone would find its path deflected, much as the sun bends a comet’s path in our universe. In 1984 Stanley Deser of Brandeis University, Roman Jackiw of the Massachusetts Institute of Technology and Gerard ’t Hooft of Utrecht University in the Netherlands worked out how quantum particles would move through such a space.

This geometry would be much simpler than the complicated pattern of curvature that gravity causes in our 4-D spacetime. Flatland would lack the equivalent of Newton’s law of attraction; instead the strength of the force would depend on objects’ velocities, and two bodies at rest would not be pulled toward each other. This simplicity is appealing. It suggests that quantizing Staruszkiewicz’s theory would be easier than quantizing full-blown general relativity in 3-D. Unfortunately, the theory is too simple: nothing is left to quantize. A 2-D space has no room for an important element of Einstein’s theory: gravitational waves.

Consider the simpler case of electromagnetism. Electric and magnetic fields are produced by electric charges and currents. As Maxwell showed, these fields can detach themselves from their sources and move freely as light waves. In the quantum version of Maxwell’s theory, the waves become photons, the quanta of light. In the same way, the gravitational fields of general relativity can detach themselves from their sources and become freely propagating gravitational waves, and physicists widely assume that a quantum theory of gravity will contain particles called gravitons that do the traveling.

A light wave has a polarization: its electric field oscillates in a direction perpendicular to its direction of motion. A gravitational wave also has a

polarization, but the pattern is more complicated: the field oscillates not in one but in two directions perpendicular to its direction of motion. Flatland has no room for this behavior. Once the direction of motion is fixed, only one perpendicular direction remains. Gravitational waves and their quantum counterparts, gravitons, simply cannot be squeezed into just two dimensions of space.

Despite occasional sparks of interest, Staruszkiewicz's discovery languished. Then, in 1989, Edward Witten of the Institute for Advanced Study in Princeton, N.J., stepped in. Witten, widely considered the world's leading mathematical physicist, had been working on a special class of fields in which waves do not propagate freely. When he realized that 2-D gravity fit into this class, he added the crucial missing ingredient: topology.

---

## How Gravity Works in 2-D

If you took 3-D space and flattened it to 2-D, matter would not just be a lot thinner. The force of gravity would behave in fundamentally different ways. Imagining gravity in 2-D has given physicists some helpful practice for how to merge Einstein's theory of gravity (the general theory of relativity) with quantum mechanics to create a quantum theory of gravity.

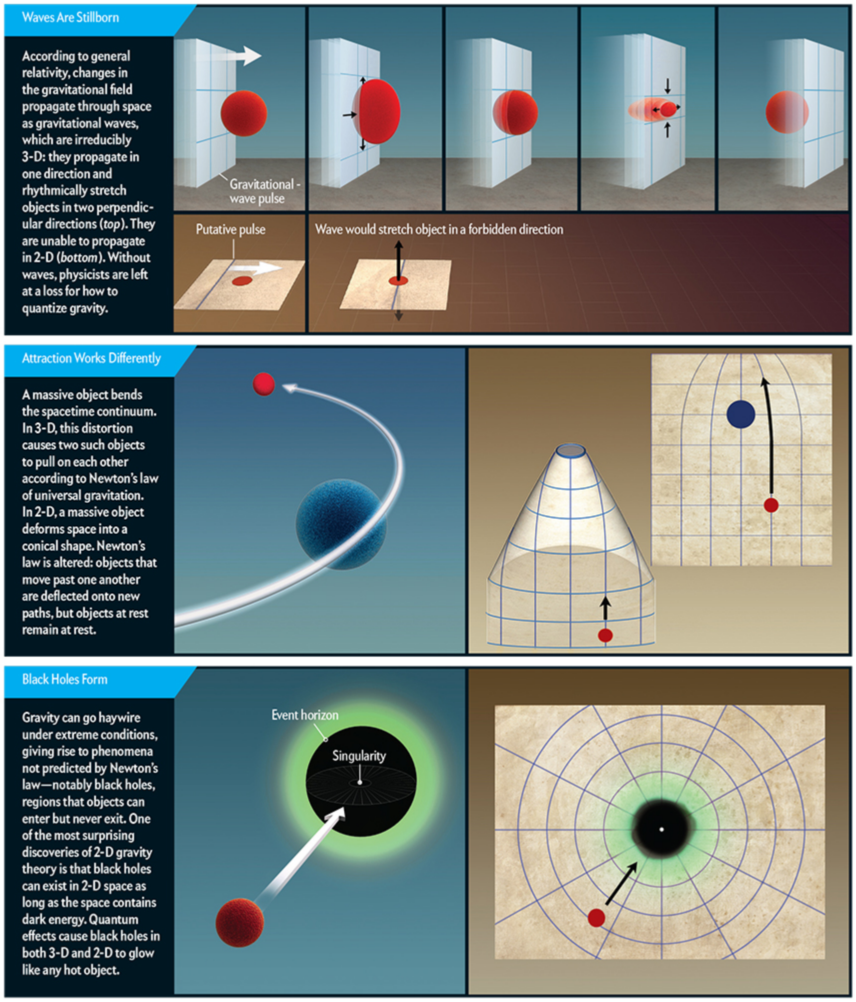



*Illustration by Malcolm Godwin*

---

## Doughnutland

Witten pointed out that even if gravity cannot propagate as waves, it can still have a dramatic effect on the overall shape of space. This effect does not arise when Flatland is just a plane; it requires a more complex topology. When an ice sculpture melts away, the details become muted, but certain features such as holes tend to last. Topology describes these features.

Two surfaces have the same topology if one can be smoothly deformed into the other without cutting, tearing or gluing. For instance, a hemisphere and a disk share the same topology: stretching the hemisphere by pulling on its perimeter yields a disk. A sphere has another topology: to turn it into a hemisphere or disk, you would need to snip out a piece. A torus, like the surface of a doughnut, has yet another. The surface of a coffee cup has the same topology as a torus: the handle looks like a torus, and the rest of the

cup can be smoothed out without cutting or tearing—hence the old mathematician's joke that a topologist can't tell a doughnut from a coffee cup.

Although tori look curved, when you consider their internal geometry rather than their shape as seen from the outside, they can actually be flat. What makes a torus a torus is the fact you can make a full loop around it in two separate directions: through the hole or around the rim. This feature will be familiar to anyone who has played any 1980s-era video game in which a combatant exiting the right side of the screen reenters on the left. The screen is flat: it obeys the rule of plane geometry, such as the fact that parallel lines never meet. Yet the topology is toroidal.

In fact, an infinite family of such tori exist—all flat but all distinct, labeled by a parameter called the modulus. What gravity in a toroidal universe does is to cause the modulus to evolve in time. A torus starts as a line at the big bang and opens up to assume an ever more square-shaped geometry as the universe expands.

Starting with Witten's results, I showed that this process could be quantized—and that doing so turns the classical theory of gravity into a quantum one. Quantum gravity in Flatland is a theory not of gravitons but of shape-shifting tori. That view is a shift from the usual picture of quantum theory as a theory of the very small. Quantum gravity in two dimensions is, in fact, a theory of the entire universe as a single object. This insight gives us a rich enough model to explore some of the fundamental conceptual problems of quantum gravity.

---

## How to Quantize Gravity in 2-D

Two-dimensional gravity has given physicists a new perspective on what gravity is. It is not necessarily a force that propagates through space—indeed, in two dimensions it cannot propagate at all. Instead gravity can also be the driver of changes in the overall shape of space. Physicists have studied a square or parallelogram universe that has been rolled into a torus. Tori of different sizes and shapes correspond to the 2-D universe at different moments in time. What happens in any small region of space mirrors the general condition of space; microcosm and macrocosm are inextricably linked.

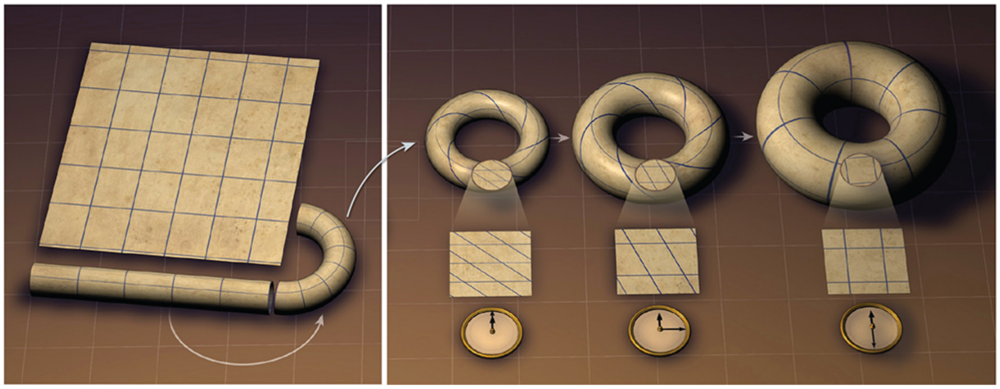


*Illustration by Malcolm Godwin*

---

## Wormholes and Bangs

In a quantum theory of gravity, unlike Einstein's theory, the topology of the universe might be able to change. That might solve some long-standing questions about the uni verse. For instance, a one-holed torus could become a two-holed one, which would amount to creating a wormhole—a backdoor passage from one location to another. Wormholes might conceivably be used as time machines. Also, the cosmos could pop out of existence or be born from sheer nothingness.

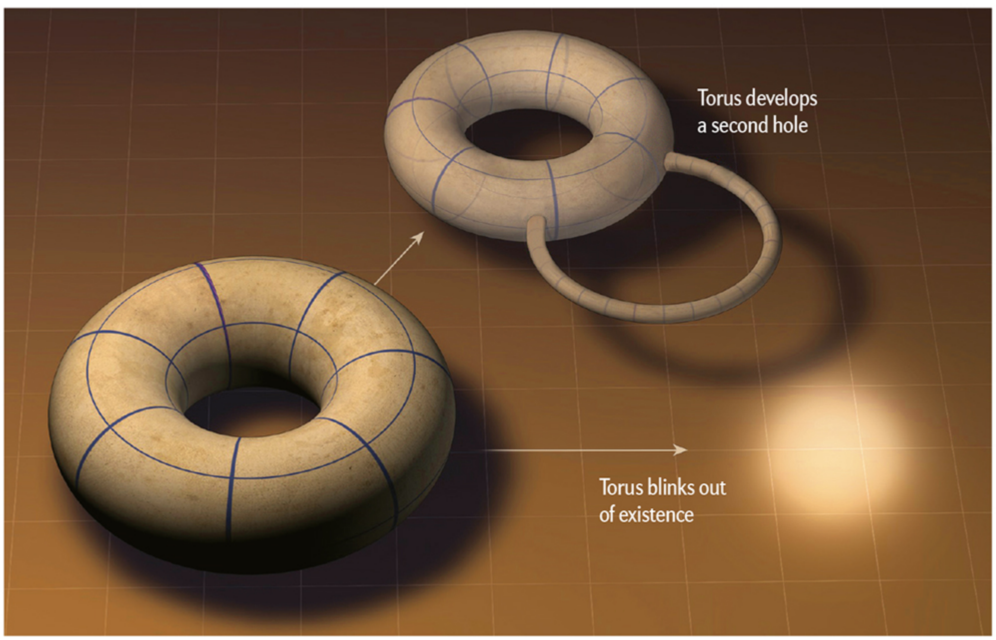



*Illustration by Malcolm Godwin*

---

## Finding the Time

Flatland gravity demonstrates, for example, how time might emerge from a fundamentally timeless reality. In one formulation of the theory, the entire universe is described by a single, quantum, wave function, similar to the mathematical device that physicists routinely use to describe atoms and subatomic particles. This wave function does not depend on time, because it

already includes all time—past, present and future—in one package. Somehow this "timeless" wave function gives rise to the change we observe in the world. The trick is to remember Einstein's aphorism that time is what is measured by a clock. Time does not stand outside the universe; it is determined by a subsystem that is correlated with the rest of the universe, just as a wall clock is correlated with Earth's rotation.

The theory offers many different clock options, and our choice defines what we mean by "time." In Doughnutland, A Square can define time by using the readings of atomic clocks in satellites, like those in the GPS. He can label time by the lengths of curves extending from the big bang, by the size of his expanding universe, or by the amount of redshift caused by its expansion. Once he has made such a choice, all other physical observables change with clock time. The modulus of the torus universe is correlated with its size, for instance, and A Square perceives this as a universe evolving in time. The theory thus bootstraps time from a timeless universe. These ideas are not new, but quantum gravity in Doughnutland has at last given us a setting in which we can do the math and check that the picture does not just look pretty but really works. Some of the definitions of time have intriguing consequences, such as implying that space can be creased.

As for the problem of observables, Doughnutland gives us a set of objectively measurable quantities—namely, the moduli. The twist is that these quantities are nonlocal: they do not reside at specific locations but describe the structure of the whole space. Anything that A Square measures is ultimately a proxy for these nonlocal quantities. In 2008 Catherine Meusburger, now at the University of Erlangen-Nürnberg in Germany, showed how these moduli relate to real cosmological measurements such as time delays and redshifts for beams of light. I have shown how they relate to objects' motion.

Flatland gravity offers good news for fans of wormholes: at least one formulation of the theory permits the topology of space to change. A Square could go to bed tonight in Sphereland and wake up tomorrow in Doughnutland, which is equivalent to creating a shortcut between two distant corners of the universe. In some versions of the theory, we can describe the creation of the universe out of nothing, the ultimate change in topology.

## On the Edge of Space

Because gravity in flatland is stunted, it used to be common knowledge among experts in the field (me included) that 2-D black holes were impossible. But in 1992 three physicists—Máximo Bañados, now at the Pontifical Catholic University of Chile in Santiago, and Claudio Bunster (then Claudio Teitelboim) and Jorge Zanelli, both at the Center for Scientific Studies in Valdivia, Chile—shocked the world, or at least our little corner of it, by showing that the theory does allow black holes, as long as the universe has a certain type of dark energy.

A so-called BTZ black hole is very much like a real black hole in our own universe. Formed from matter collapsing under its own weight, it is surrounded by an event horizon, a one-way barrier from which nothing can escape. To an observer who remains on the outside, the event horizon looks like an edge of the universe: any object that falls through the horizon is completely cut off from us. Per Hawking's calculations, A Square should see it glow at a temperature that depends on its mass and spin.

That result presents a puzzle. Lacking gravitational waves or gravitons, Flatland gravity should also lack the gravitational degrees of freedom that would explain black hole temperature. Yet they sneak in anyway. The reason is that the event horizon itself provides some additional structure that empty 2-D space lacks. The horizon exists at a certain location, which, mathematically, augments the raw theory with some additional quantities. Vibrations that wiggle the horizon provide degrees of freedom. Remarkably, we find that they exactly reproduce Hawking's results.

Because the degrees of freedom are features of the horizon, they reside, in a sense, on the edge of Flatland itself. So they are a concrete realization of a fascinating proposal about the nature of quantum gravity, the holographic principle. This principle suggests that dimension may be a fungible concept. Just as a hologram captures a three-dimensional image on a flat 2-D film, many physicists speculate that the physics of a d-dimensional world can be completely captured by a simpler theory in d–1 dimensions. In string theory—a leading effort to unify general relativity and quantum mechanics—this idea led in the late 1990s to a novel approach for creating a quantum theory of gravity.

Flatland gravity provides a simplified scenario to test that approach. In 2007 Witten and Alexander Maloney, now at McGill University, again surprised the physics world by suggesting that the holographic predictions appear to fail for the simplest form of 2-D gravity. The theory, they found, seemed to predict impossible thermal properties for black holes. This unexpected result suggests that gravity is an even more subtle phenomenon than we had suspected, and the response has been a fresh surge of Flatland research.

Perhaps gravity simply does not make sense by itself but must work in partnership with other kinds of forces and particles. Perhaps Einstein's theory needs to be revised. Perhaps we need to find a way to put back some local degrees of freedom. Perhaps the holographic principle does not always hold. Or perhaps space, like time, is not a fundamental ingredient of the universe. Whatever the answer, Flatland gravity has pointed us in a direction we might not otherwise have taken.

Although we cannot make a real 2-D black hole, we might be able to test some of the predictions of the Flatland model experimentally. Several laboratories around the world are working on 2-D analogues of black holes. For example, a fluid flowing faster than the speed of sound produces a sonic event horizon from which sound waves cannot escape. Experimenters have also built 2-D black holes by using electromagnetic waves that are confined to surfaces. Such analogues should also exhibit a quantum glow in much the same way a black hole does.

Quantum gravity in Flatland began as a playground for physicists, a simple setting in which to explore ideas about real-world quantum gravity. It has already taught us valuable lessons about time, observables and topology that are carrying over to real 3-D gravity. The model has surprised us with its richness: the unexpectedly important role of topology, its remarkable black holes, its strange holographic properties. Perhaps soon we will fully understand what it is like to be a square living in a flat world.

---

## Flatland for Real

A laboratory system that mimics Flatland, developed by Igor I. Smolyaninov and his colleagues at the University of Maryland, is a metal surface along which electro magnetic waves propagate. These 2-D analogues of light are known as surface plasmons. A liquid droplet traps them much as a 3-D black hole traps photons; the analogue of the event horizon shows up as a white rim below right). Just as theorists find Flatland gravity a

useful warm-up exercise for unifying physics, experimentalists think that 2-D systems will have practical applications in optics.

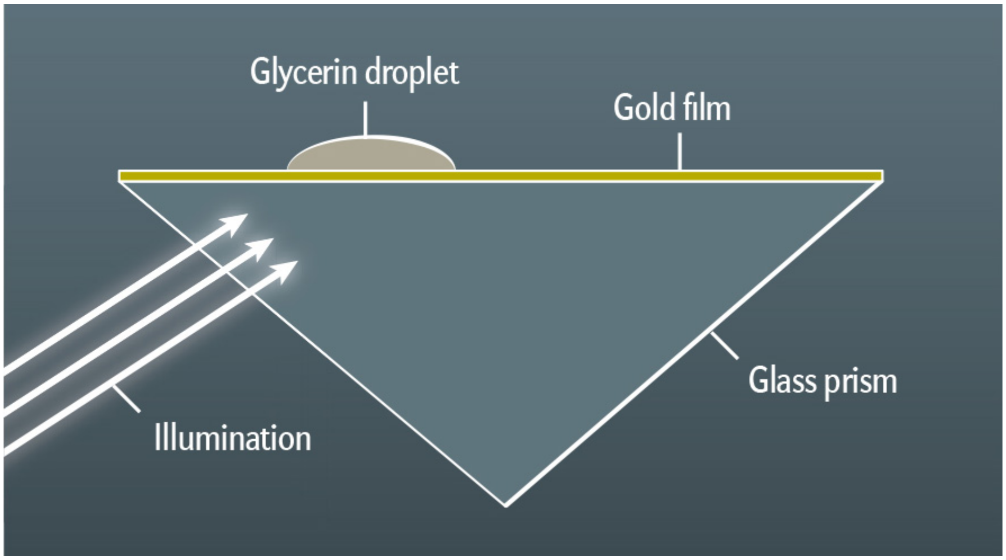



*Illustration by Jen Christiansen*

---

--Originally published: Scientific American 24(4); 76-83 (Winter 2015).

# What Is Real?

by Meinard Kuhlman

Physicists routinely describe the universe as being made of tiny subatomic particles that push and pull on one another by means of force fields. They call their subject "particle physics" and their instruments "particle accelerators." They hew to a Legolike model of the world. But this view sweeps a little-known fact under the rug: the particle interpretation of quantum physics, as well as the field interpretation, stretches our conventional notions of "particle" and "field" to such an extent that ever more people think the world might be made of something else entirely.

The problem is not that physicists lack a valid theory of the subatomic realm. They do have one: it is called quantum field theory. Theorists developed it between the late 1920s and early 1950s by merging the earlier theory of quantum mechanics with Einstein's special theory of relativity. Quantum field theory provides the conceptual underpinnings of the Standard Model of particle physics, which describes the fundamental building blocks of matter and their interactions in one common framework. In terms of empirical precision, it is the most successful theory in the history of science. Physicists use it every day to calculate the aftermath of particle collisions, the synthesis of matter in the big bang, the extreme conditions inside atomic nuclei, and much besides.

So it may come as a surprise that physicists are not even sure what the theory says—what its "ontology," or basic physical picture, is. This confusion is separate from the much discussed mysteries of quantum mechanics, such as whether a cat in a sealed box can be both alive and dead at the same time.

The unsettled interpretation of quantum field theory is hobbling progress toward probing whatever physics lies beyond the Standard Model, such as

string theory. It is perilous to formulate a new theory when we do not understand the theory we already have. At first glance, the content of the Standard Model appears obvious. It consists, first, of groups of elementary particles, such as quarks and electrons, and, second, of four types of force fields, which mediate the interactions among those particles. This picture appears on classroom walls and in *Scientific American* articles. Yet however compelling it might appear, it is not at all satisfactory.

For starters, the two categories blur together. Quantum field theory assigns a field to each type of elementary particle, so there is an electron field as surely as there is an electron. At the same time, the force fields are quantized rather than continuous, which gives rise to particles such as the photon. So the distinction between particles and fields appears to be artificial, and physicists often speak as if one or the other is more fundamental.

Debate has swirled over this point—over whether quantum field theory is ultimately about particles or about fields. It started as a battle of titans, with eminent physicists and philosophers on both sides. Even today both concepts are still in use for illustrative purposes, although most physicists would admit that the classical conceptions do not match what the theory actually describes. If the mental images conjured up by the words "particle" and "field" do not match what the theory says, physicists and philosophers must figure out what to put in their place.

With the two standard, classical options gridlocked, some philosophers of physics have been formulating more radical alternatives. They suggest that the most basic constituents of the material world are intangible entities such as relations or properties. One particularly radical idea is that everything can be reduced to intangibles alone, without any reference to individual things. It is a counterintuitive and revolutionary idea, but some argue that physics is forcing it on us.

## The Trouble with Particles

When most people, including experts, think of subatomic reality, they imagine particles that behave like little billiard balls rebounding o one another. But this notion of particles is a holdover of a worldview that dates to the ancient Greek atomists—a view that reached its pinnacle in the

theories of Isaac Newton. Several overlapping lines of thought make it clear that the core units of quantum field theory do not behave like billiard balls at all.

First, the classical concept of a particle implies something that exists in a certain location. But the “particles” of quantum field theory do not have well-defined locations: a particle inside your body is not strictly inside your body. An observer attempting to measure its position has a small but nonzero probability of detecting it in the most remote places of the universe. This contradiction was evident in the earliest formulations of quantum mechanics but became worse when theorists merged quantum mechanics with relativity theory. Relativistic quantum particles are extremely slippery; they do not reside in any specific region of the universe at all.

Second, let us suppose you had a particle localized in your kitchen. Your friend, looking at your house from a passing car, might see the particle spread out over the entire universe. What is localized for you is delocalized for your friend. Not only does the location of the particle depend on your point of view, so does the fact that the particle *has* a location. In this case, it does not make sense to assume localized particles as the basic entities.

Third, even if you give up trying to pinpoint particles and simply count them, you are in trouble. Suppose you want to know the number of particles in your house. You go around the house and find three particles in the dining room, five under the bed, eight in a kitchen cabinet, and so on. Now add them up. To your dismay, the sum will not be the total number of particles. That number in quantum field theory is a property of the house as a whole; to determine it, you would have to do the impossible and measure the whole house in one go, rather than room by room.

An extreme case of particles’ being unpinpointable is the vacuum, which has paradoxical properties in quantum field theory. Look closely at any finite region of an overall vacuum—by definition, a zero-particle state—and you may observe something very different from a vacuum. In other words, your house can be totally empty even though you find particles all over the place.

Another striking feature of the vacuum in quantum field theory is known as the Unruh effect. An astronaut at rest may think he or she is in a vacuum, whereas an astronaut in an accelerating spaceship will feel immersed in a thermal bath of innumerable particles. This discrepancy between viewpoints also occurs at the perimeter of black holes and leads to paradoxical conclusions about the fate of infalling matter.

If a vacuum filled with particles sounds absurd, that is because the classical notion of a particle is misleading us; what the theory is describing must really be something else. If the number of particles is observer-dependent, then it seems incoherent to assume that particles are basic. We can accept many features to be observer-dependent—but not the very fact of how many basic building blocks there are.

Finally, the theory dictates that particles can lose their individuality. In the puzzling phenomenon of quantum entanglement, particles can become assimilated into a larger system and give up the properties that distinguish them from one another. The presumptive particles share not only innate features such as mass and charge but also spatial and temporal properties such as the range of positions over which they might be found. When particles are entangled, an observer has no way of telling one from the other. At that point, do you really have two objects anymore?

A theorist might simply decree that our would-be two particles are two distinct individuals. Philosophers call this diktat “primitive thisness.” By definition, this thisness is unobservable.

Most physicists and philosophers are very skeptical of such ad hoc moves. Rather, it seems, you no longer have two particles anymore. The entangled system behaves as an indivisible whole, and the notion of a part, let alone a particle, loses its meaning.

These theoretical problems with particles fly in the face of experience. What do “particle detectors” detect if not particles? The answer is: we don’t see particles— we infer them. All a detector registers is a large number of separate excitations of the sensor material. We run into trouble when we connect the dots and infer the existence of particles having trajectories that can be traced in time.

(Caveat: Some minority interpretations of quantum physics do assume well-defined trajectories. But they suffer from their own difficulties, and I stick to the standard view.)

So let us take stock. We think of particles as tiny billiard balls, but the things that modern physicists call “particles” are nothing like that. According to quantum field theory, objects cannot be localized in any finite region of space, no matter how large or fuzzy it is. Moreover, the number of the putative particles depends on the state of motion of the observer. All these results taken together sound the death knell for the idea that nature is composed of anything akin to ball-like particles.

On the basis of these and other insights, one must conclude that “particle physics” is a misnomer: despite the fact that physicists keep talking about particles, there are no such things. One may adopt the phrase “quantum particle,” but what justifies the use of the word “particle” if almost nothing of the classical notion of particles has survived? It is better to bite the bullet and abandon the concept altogether.

Some see these difficulties as indirect evidence that quantum field theory describes only fields. By this reasoning, particles are ripples in a field that fills space like an invisible fluid. Yet as we will see now, quantum field theory cannot be readily interpreted in terms of fields, either.

---

## Not Just Tiny Billiard Balls

You would be forgiven for thinking that particle physics is about particles. It turns out, though, that the “particles” described by quantum theory do not fit the usual sense of the term, which refers to discrete, localized building blocks of matter. They lack, for instance, the four classical attributes listed below.

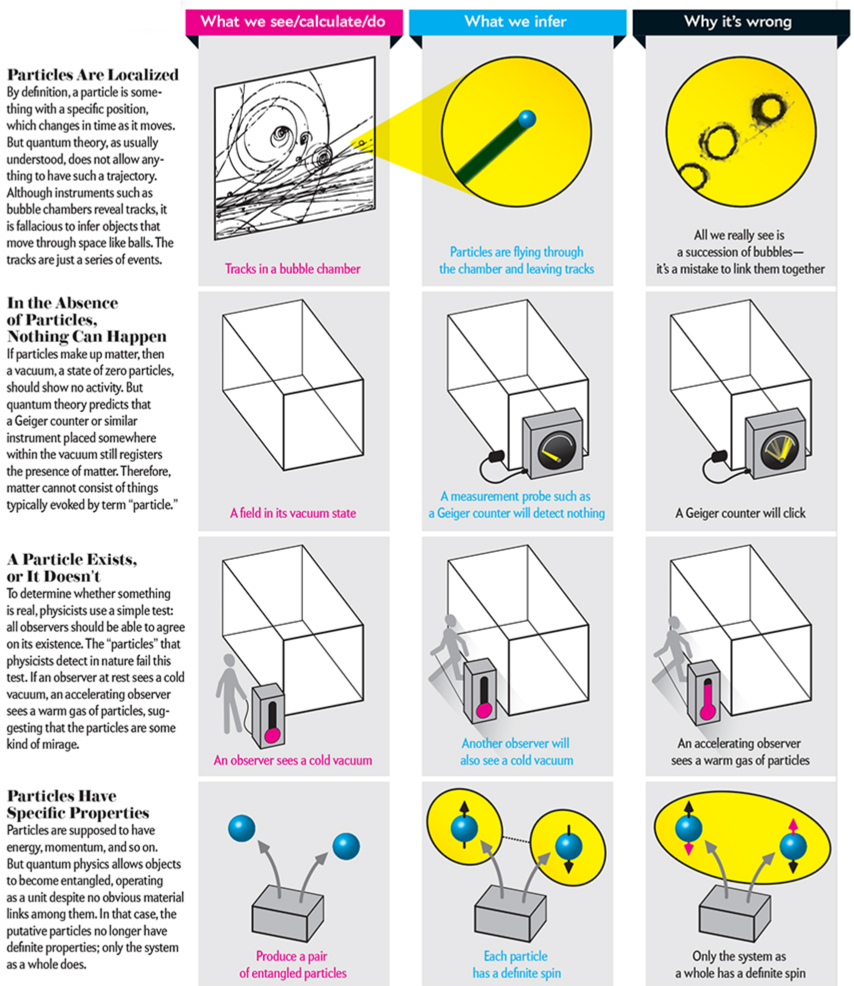



*Illustration by Jen Christiansen.*

---

## The Trouble with Fields

The name "quantum field theory" naturally connotes a theory that deals with quantum versions of classical fields, such as the electric and magnetic fields. But what is a "quantum version"? The term "field" conjures up magnetic fields that cause iron filings to align themselves around a bar magnet and electric fields that cause hair to stand up on end, but a quantum field is so different from a classical one that even theoretical physicists admit they can barely visualize it.

Classically, a field assigns a physical quantity, such as temperature or electric field strength, to each point in spacetime. A quantum field instead assigns abstract mathematical entities, which represent the type of measurements you could conduct, rather than the result you would obtain. Some mathematical constructions in the theory do represent physical

values, but these cannot be assigned to points in spacetime, only to smeared-out regions.

Physicists originally developed quantum field theory by “quantizing” classical field theory. In this procedure, theorists go through an equation and replace physical values with “operators,” which are mathematical operations such as differentiation or taking the square root. Some operators can correspond to specific physical processes such as the emission and absorption of light. Operators place a layer of abstraction between the theory and reality.

A classical field is like a weather map that shows the temperature in various cities. The quantum version is like a weather map that does not show you “40 degrees,” but rather “√.” To obtain an actual temperature value, you would need to go through another step of applying the operator to a mathematical entity, known as a state vector, that represents the configuration of the system in question.

On some level, this peculiarity of quantum fields does not seem surprising. Quantum mechanics—the theory on which quantum field theory is based—does not traffic in determinate values either but only in probabilities. The situation, however, seems weirder in quantum field theory because the supposedly fundamental entities, the quantum fields, do not even specify any probabilities; for that, they must be combined with the state vector.

The need to apply the quantum field to the state vector makes the theory very difficult to interpret, to translate into something physical you can imagine and manipulate in your mind. The state vector is holistic; it describes the system as a whole and does not refer to any particular location. Its role undermines the defining feature of fields, which is that they are spread out over spacetime. A classical field lets you envision phenomena such as light as propagation of waves across space. The quantum field takes away this picture and leaves us at a loss to say how the world works.

Clearly, then, the standard picture of elementary particles and mediating force fields is not a satisfactory ontology of the physical world. It is not at all clear what a particle or field even is. A common response is that

particles and fields should be seen as complementary aspects of reality. But that characterization does not help, because neither of these conceptions works even in those cases where we are supposed to see one or the other aspect in purity. Fortunately, the particle and field views do not exhaust the possible physical interpretations of quantum field theory.

---

### No Field of Dreams

Physicists call their leading theory of matter "quantum field theory." That sounds like a theory of fields. Yet the fields supposedly described by the theory are not what physicists classically understand by the term "field."

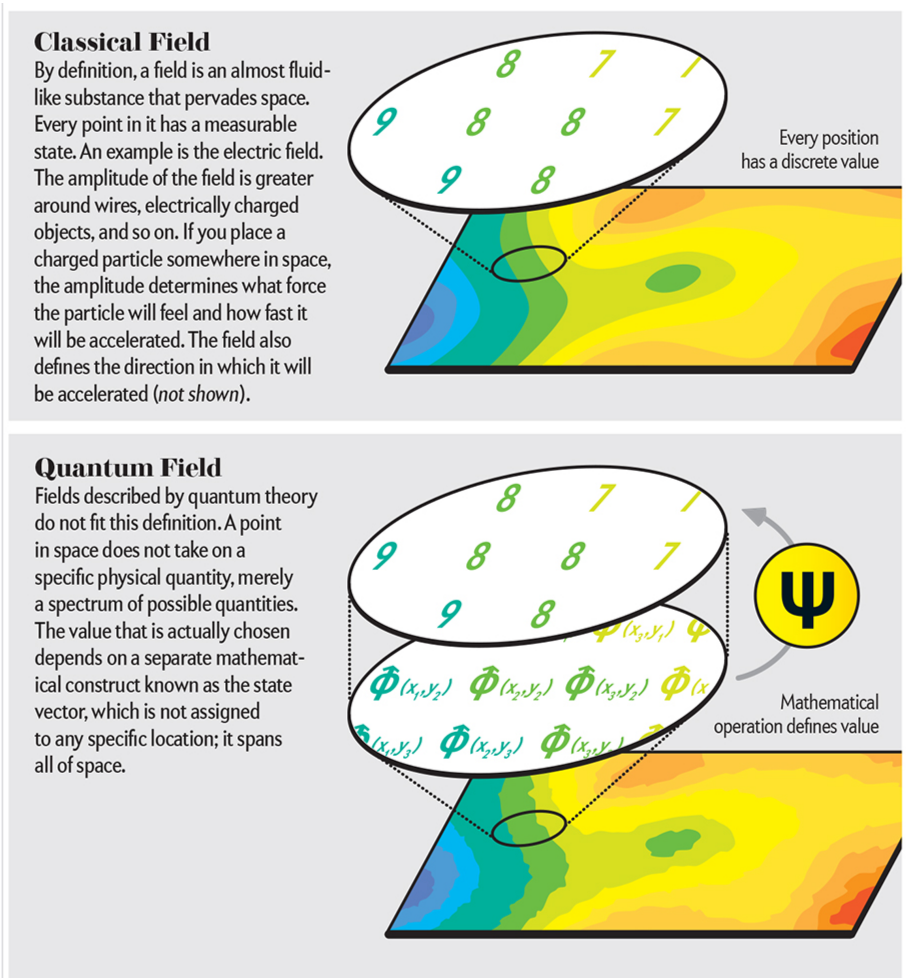



*Illustration by Jen Christiansen.*

---

## Structures to the Rescue?

A growing number of people think that what really matters are not things but the relations in which those things stand. Such a view breaks with traditional atomistic or pointillist conceptions of the material world in a

more radical way than even the severest modifications of particle and field ontologies could do.

Initially this position, known as structural realism, came in a fairly moderate version known as epistemic structural realism. It runs as follows: we may never know the real natures of things but only how they are related to one another.

Consider mass. Do you ever see mass itself? No. You see only what it means for other entities or, concretely, how one massive body is related to another massive body through the local gravitational field.

The structure of the world, reflecting how things are interrelated, is the most enduring part of physics theories. New theories may overturn our conception of the basic building blocks of the world, but they tend to preserve the structures. That is how scientists can make progress.

Now a question arises: What is the reason that we can know only the relations among things and not the things themselves? The straightforward answer is that relations are all there is. This leap makes structural realism a more radical proposition, called ontic structural realism.

The myriad symmetries of modern physics lend support to ontic structural realism. In quantum mechanics as well as in Einstein’s theory of gravitation, certain changes in the configuration of the world—known as symmetry transformations—have no empirical consequences. These transformations exchange the individual things that make up the world but leave their relations the same.

By analogy, consider a mirror-symmetric face. A mirror swaps the left eye for the right eye, the left nostril for the right, and so on. Yet all the relative positions of facial features remain. Those relations are what truly define a face, whereas labels such as “left” and “right” depend on your vantage point. The things we have been calling “particles” and “fields” possess more abstract symmetries, but the idea is the same.

By the principle of Occam’s razor, physicists and philosophers prefer ideas that can explain the same phenomena with the fewest assumptions. In this case, you can construct a perfectly valid theory by positing the existence of specific relations without additionally assuming individual

things. So proponents of ontic structural realism say we might as well dispense with things and assume that the world is made of structures, or nets of relations.

In everyday life we encounter many situations where only relations count and where it would be distracting to describe the things that are related. In a subway network, for example, it is crucial to know how the different stations are connected. In London, St. Paul’s is directly connected to Holborn, whereas from Blackfriars you need to change lines at least once, even though Blackfriars is closer to Holborn than St. Paul’s. It is the structure of the connections that matters primarily. The fact that Blackfriars Tube station has been renovated into a nice new station does not matter to someone trying to navigate the system.

Other examples of structures that take priority over their material realization are the World Wide Web, the brain’s neural network and the genome. All of them still function even when individual computers, cells, atoms and people die. These examples are loose analogies, although they are close in spirit to the technical arguments that apply to quantum field theory.

A closely related line of reasoning exploits quantum entanglement to make the case that structures are the basis of reality. The entanglement of two quantum particles is a holistic effect. All the intrinsic properties of the two particles, such as electric charge, together with all their extrinsic properties, such as position, still do not determine the state of the two-particle system. The whole is more than the sum of its parts. The atomistic picture of the world, in which everything is determined by the properties of the most elementary building blocks and how they are related in spacetime, breaks down. Instead of considering particles primary and entanglement secondary, perhaps we should think about it the other way round.

You may find it is strange that there could be relations without relata—without any objects that stand in that relation. It sounds like having a marriage without spouses. You are not alone. Many physicists and philosophers find it bizarre, too, thinking it impossible to get solid objects merely on the basis of relations.

Some proponents of ontic structural realism try to compromise. They do not deny objects exist; they merely claim that relations, or structures, are more fundamental. In other words, the properties of objects are not intrinsic—they come only from their relations with other things.

But this position seems wishy-washy. Anyone would agree that objects have relations. The only interesting and new position would be that everything emerges purely on the basis of relations. All in all, structural realism is a provocative idea, but it needs further development before we will know whether it can rescue us from our interpretive trouble.

**Bundles of Properties**

A second alternative for the meaning of quantum field theory starts from a simple insight. Although the particle and field interpretations are traditionally considered to be radically different from each other, they have something crucial in common. Both assume that the fundamental items of the material world are persistent individual entities to which properties can be ascribed. These entities are either particles or, in the case of field theory, spacetime points. Many philosophers, including me, think this division into objects and properties may be the deep reason that the particle and field approaches both run into difficulties. We think it would be better to view properties as the one and only fundamental category.

Traditionally, people assume that properties are "universals"—in other words, that they belong to an abstract, general category. Properties are always possessed by particular things; they cannot exist independently. (To be sure, Plato did think of them as existing independently—but only in some higher realm, not in the world that exists in space and time.) When you think of red, for example, you usually think of particular red things and not of some freely floating item called "redness."

But you could invert this way of thinking. You can regard properties as having an existence that is independent of objects that possess them. Properties may be what philosophers call "particulars"—concrete, individual entities. What we commonly call a thing may be just a bundle of properties: color, shape, consistency, and so on.

Because this conception of properties as particulars rather than universals differs from the traditional view, philosophers have introduced a new term

to describe them: “tropes.” It sounds a bit funny, and unfortunately the term brings inappropriate connotations with it, but it is established by now.

Construing things as bundles of properties is not how we usually conceptualize the world, but it becomes less mysterious if we try to unlearn how we usually think about the world and set ourselves back to the very first years of life. As infants, when we see and experience a ball for the first time, we do not actually perceive a ball, strictly speaking. What we perceive is a round shape, some shade of red, with a certain elastic touch. Only later we do associate this bundle of perceptions with a coherent object of a certain kind—namely, a ball. Next time we see a ball, we essentially say, “Look, a ball,” and forget how much conceptual apparatus is involved in this seemingly immediate perception.

In trope ontology, we return to the direct perceptions of infancy. Out there in the world, things are nothing but bundles of properties. It is not that we first have a ball and then attach properties to it. Rather we have properties and call it a ball. There is nothing to a ball but its properties. Applying this idea to quantum field theory, what we call an electron is in fact a bundle of properties or tropes: three fixed, essential properties (mass, charge and spin), plus numerous changing, nonessential properties (position and velocity).

The concept of tropes helps make sense of quantum field theory. For instance, the theory predicts a particularly mind-boggling behavior of the vacuum: the average value of the number of particles is zero, yet the vacuum seethes with activity. Countless processes take place all the time, involving the creation and subsequent destruction of all kinds of elementary particles.

In a particle ontology, this activity is paradoxical. If particles are fundamental, then how can they materialize? What do they materialize out of?

In the trope ontology, the situation is natural. The vacuum, though empty of particles, contains properties. A particle is what you get when those properties bundle themselves together in a certain way.

## Physics and Metaphysics

How can there be so much fundamental controversy about a theory that is as empirically successful as quantum field theory? The answer is straightforward. Although the theory tells us what we can measure, it speaks in riddles when it comes to the nature of what ever entities give rise to our observations. The theory accounts for our observations in terms of quarks, muons, photons and sundry quantum fields, but it does not tell us what a photon or a quantum field really is. And it does not need to, because theories of physics can be empirically valid largely without settling such metaphysical questions.

For many physicists, that is enough. They adopt a so-called instrumentalist attitude: they deny that scientific theories are meant to represent the world in the first place. For them, theories are only instruments for making experimental predictions. Still, most scientists have the strong intuition that their theories do depict at least some aspects of nature as it is before we make a measurement. After all, why else do science, if not to understand the world?

Acquiring a comprehensive picture of the physical world requires the combination of physics with philosophy. The two disciplines are complementary. Metaphysics supplies various competing frameworks for the ontology of the material world, although beyond questions of internal consistency, it cannot decide among them. Physics, for its part, lacks a coherent account of fundamental issues, such as the definition of objects, the role of individuality, the status of properties, the relation of things and properties, and the significance of space and time.

The union of the two disciplines is especially important at times when physicists find themselves revisiting the very foundations of their subject. Metaphysical thinking guided Isaac Newton and Albert Einstein, and it is influencing many of those who are trying to unify quantum field theory with Einstein’s theory of gravitation. Philosophers have written libraries full of books and papers about quantum mechanics and gravity theory, whereas we are only beginning to explore the reality embodied in quantum field theory. The alternatives to the standard particle and field views that we are developing may inspire physicists in their struggle to achieve the grand unification.

--Originally published: Scientific American 24(4); 84-91 (Winter 2015).

# Quantum Weirdness? It's All in Your Mind

by Hans Christian von Baeyer

Flawlessly accounting for the behavior of matter on scales from the subatomic to the astronomical, quantum mechanics is the most successful theory in all the physical sciences. It is also the weirdest.

In the quantum realm, particles seem to be in two places at once, information appears to travel faster than the speed of light, and cats can be dead and alive at the same time. Physicists have grappled with the quantum world's apparent paradoxes for nine decades, with little to show for their struggles. Unlike evolution and cosmology, whose truths have been incorporated into the general intellectual landscape, quantum theory is still considered (even by many physicists) to be a bizarre anomaly, a powerful recipe book for building gadgets but good for little else. The deep confusion about the meaning of quantum theory will continue to add fuel to the perception that the profound things it is so urgently trying to tell us about our world are irrelevant to everyday life and too weird to matter.

In 2001 a team of researchers began to develop a model that either eliminates the quantum paradoxes or puts them in a less troubling form. The model, known as Quantum Bayesianism, or QBism for short, reimagines the entity that lies at the heart of quantum weirdness—the wave function.

In the conventional view of quantum theory, all the properties of any isolated system, such as an atom, are encapsulated mathematically by the system's wave function. If you want to predict how likely it is that an electron in that atom will appear at a certain spot, for example, you can calculate that probability from its wave function. This mathematical construct is a tremendously useful tool for both theoretical and experimental physics. But problems arise when physicists assume that a wave function is real.

QBism, which combines quantum theory with probability theory, maintains that the wave function has no objective reality. Instead QBism portrays the wave function as a mathematical guidebook. An observer can use it to anticipate how things behave in the quantum world.

Specifically, the observer employs the wave function to assign his or her belief that properties of a quantum system will have particular values, realizing that one's own actions affect the system and change those properties in inherently uncertain ways. Another observer, using a wave function that describes the world as that person sees it, may come to a completely different conclusion about the same quantum system. One system—one event—can have as many different wave functions as there are observers. After they have communicated with one another and modified their private wave functions to account for the newly acquired knowledge, a coherent worldview emerges.

Seen this way, the wave function "may well be the most powerful abstraction we have ever found," says theoretical physicist N. David Mermin of Cornell University, a recent convert to QBism.

### The Unreal Quantum

The notion that the wave function isn't real dates back to the 1930s and the writings of Niels Bohr, one of the founding fathers of quantum mechanics. He considered it part of quantum theory's "purely symbolic" formalism—a computational tool, no more. QBism is the first model to give mathematical backbone to Bohr's assertion. It melds quantum theory with Bayesian statistics, a 200-year-old discipline that defines "probability" as something like "subjective belief." Bayesian statistics also gives formal mathematical rules for how to update one's subjective beliefs in light of new information. By interpreting the wave function as a subjective belief and subject to revision by the rules of Bayesian statistics, the mysterious paradoxes of quantum mechanics vanish, QBism's proponents say.

Consider again an electron in an atom. We set up an experiment to detect the particle, and we find it in one particular location. But as soon as we stop looking, the electron's wave function spreads out. That seems to imply that the electron could be in many different places at once. Yet whenever we detect the particle again, we always find it occupying just one position.

According to the standard way of thinking, the act of observation causes the wave function to instantaneously “collapse,” snapping the electron into a particular location.

Because the collapse happens everywhere at exactly the same time, it seems to violate the principle of locality—the idea that any change in an object must be caused by another object in its immediate surroundings. This, in turn, leads to some of the puzzles that Albert Einstein called “spooky action at a distance.”

From the very birth of quantum mechanics, physicists saw the collapse of the wave function as a paradoxical and deeply disturbing feature of the theory. Its uneasy mysteries pushed physicists to develop alternative versions of quantum mechanics, with mixed success.

Yet QBism says that there is no paradox. The wave function’s collapse is just an observer suddenly and discontinuously revising probability assignments based on new information, in the same way that a doctor would revise a cancer patient’s prognosis based on a new CT scan. The quantum system hasn’t undergone some strange and inexplicable change; the change is in the wave function, which is chosen by the observer to encapsulate the person’s expectations.

We can apply this way of thinking to the famous paradox of Schrödinger’s cat. Quantum physicist Erwin Schrödinger imagined a sealed box with a live cat, a vial of poison and a radioactive atom. The atom has a 50–50 chance of decaying within an hour, according to the rules of quantum mechanics. If the atom decays, a hammer will smash the vial and release the poison, killing the cat. If it doesn’t, the cat lives.

Now run the experiment—but don’t look inside the box. After an hour has gone by, traditional quantum theory would hold that the atom’s wave function is in a superposition of two states: decayed and not decayed. But because you haven’t yet observed what is inside the box, the superposition extends further. The hammer is also in a superposition, as is the vial of poison. And most grotesquely, the standard quantum-mechanical formalism implies that the cat is in a superposition—it is both alive and dead at the same time.

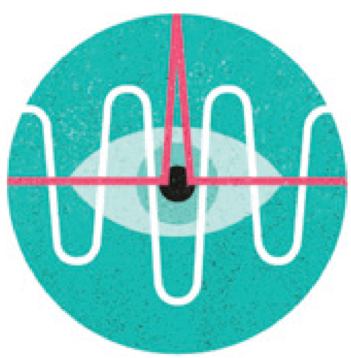
By insisting that the wave function is a subjective property of the observer, rather than an objective property of the cat in the box, QBism eliminates the puzzle. Common sense says that of *course* the cat is either alive or dead (and not both). Sure, the wave function of this system represents a superposition of "alive" and "dead," but a wave function is just a description of the observer's beliefs. Asserting that the cat is truly both alive and dead is akin to a baseball fan saying that the Yankees are stuck in a superposition of both "won" and "lost" until the person sees the box score. It's an absurdity, a megalomaniac's delusion that one's personal state of mind makes the world come into being.

The hope is that by removing the paradoxes, QBism will help physicists home in on the truly fundamental features of quantum theory—whatever they turn out to be—and "prevent them from wasting their time asking silly questions about illusory puzzles," Mermin says.

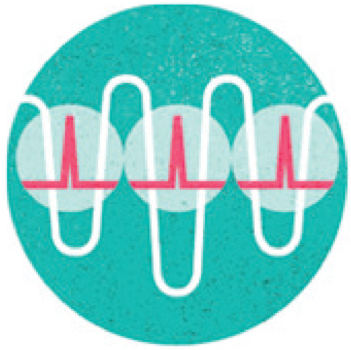
---

## Four Interpretations of Quantum Mechanics

What is really happening in the quantum world? Scientists have offered about a dozen different interpretations of what the mathematical formalism implies. Quantum Bayesianism is perhaps the most radical; these four alternatives are among the most popular

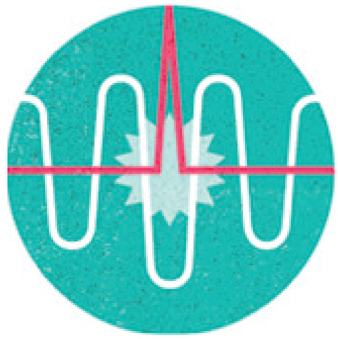


**THE COPENHAGEN INTERPRETATION**, developed principally by Niels Bohr and Werner Heisenberg at the former's institute in Copenhagen, is the orthodox version of quantum mechanics. The measurable properties of a system such as an atom are collectively called its quantum state. The quantum state, in turn, is described by either a matrix, which resembles a spreadsheet, or a formula called the wave function, which represents a map of possibilities. Contact with the real world is made by the Born rule, a recipe for obtaining measurable probabilities for a given quantum state (and for which Heisen berg's mentor Max Born received a Nobel Prize). During a measurement an observer causes a collapse of the quantum state into a new state that describes the actual outcome of the experiment. The instantaneous collapse implies that actions can have effects that travel faster than light.



**THE MANY-WORLDS INTERPRETATION**. The most direct way of avoiding the conundrum of quantum state collapse is to eliminate it. This drastic move has gathered many supporters in recent years. The many-worlds interpretation posits a single quantum state of the world, which unfolds smoothly and predictably. When an experiment is performed to ascertain which of two slits an electron traversed, for example, the quantum state does not collapse onto one slit. Instead the world actually splits into two branches. We, the observers of the real world, reside on one branch and are unaware of the other. Thus, the universe really branches out like a tree into a vast multiverse in which every possible outcome actually occurs in one of an infinity of distinct, real universes. The principal drawbacks of this interpretation, aside from its exorbitant demands on our imagination, are its failure to account for the "measurements" that lead to branching and its difficulty in justifying the Born rule

**THE GUIDING FIELD INTERPRETATION**. A number of physicists, including Albert Einstein for a while, favored rewriting the mathematical apparatus of quantum mechanics to include a real physical field of force that controls the motion of a particle. Unfortunately, this appealing image breaks down as soon as several particles, say N of them, are involved. They do not move in our familiar three-dimensional space but in an abstract space with 3 N dimensions. More troubling is the fact that the guiding field exerts an action-at-a-distance force, in which physical effects are transmitted instantaneously over large distances.



**SPONTANEOUS COLLAPSE THEORIES**. Rather than eliminating the observer-triggered collapse, these theories posit that collapses are entirely natural—they happen spontaneously, though rarely, to every quantum system but become significant when the quantum system interacts with a macroscopic object. Yet they require the invention of an entirely new mechanism of collapse. As long as the collapse mechanism cannot be tested experimentally, it constitutes a new assumption that is every bit as mysterious as the observer-induced collapse it is designed to replace.

*Illustration by Anna-Kaisa Jormanainen*

---

## The Fix for Quantum Absurdity

One difference between QBism and the standard interpretation of quantum mechanics can be shown by Schrödinger's thought experiment of a cat sealed in a box with a vial of poison. The system is rigged to have a 50–50 chance that a quantum event will occur, break the vial and kill the cat. As long as the outcome is unknown, the wave function describing the system is in a superposition of states. The standard interpretation considers the cat simultaneously "alive" and "dead" until an observation collapses the wave function and seals the cat's fate. In the QBist view, the wave function merely describes the limited information about the system available to the observer; the superposition applies only to this information. Opening the box adds new information that resolves the uncertainty.

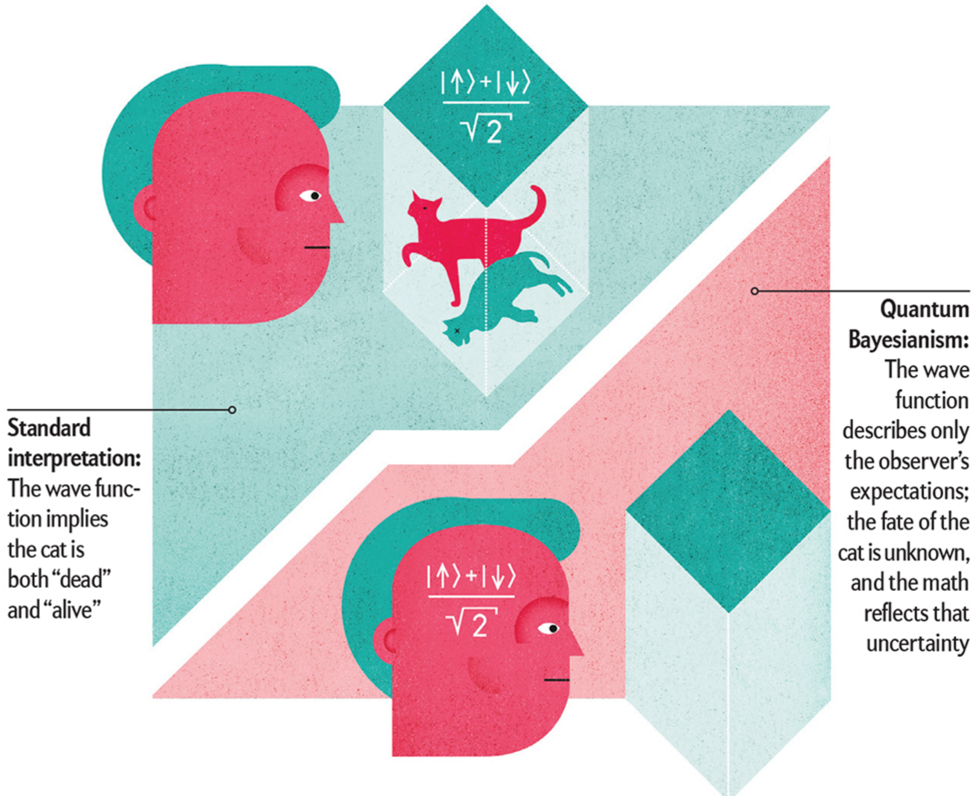



*Illustration by Anna-Kaisa Jormanainen*

## The Troublemaker

Qbism was born in a short paper, the preprint published in November 2001 under the title "Quantum Probabilities as Bayesian Probabilities," by Carlton M. Caves of the University of New Mexico, Christopher A. Fuchs, then at Bell Labs in Murray Hill, N.J., and Rüdiger Schack of the University of London. All three are experienced quantum information theorists, and their respective affiliations with a physics department, an industrial laboratory and a department of mathematics illustrate the interdisciplinary nature of their field.

Since then, Fuchs has moved to the University of Massachusetts Boston and assumed the role of QBism's chief spokesperson. He is a compact Texan with a cheerful disposition. A sandy-colored cow lick at his hairline hints at his irrepressible, irreverent sense of humor. Colleagues are not surprised when he opens an article with the words "In this paper, I try to cause some good-natured trouble."

The core of Fuchs's style is the conviction that science is quintessentially a communal activity and that profound insight is won only through vigorous intellectual combat. He is a whirlwind of activity, lugging his laptop around the world in a beat-up backpack, organizing conferences, chairing scientific sessions and giving lectures at universities.

In this spirit, Fuchs has pioneered a new form of literature. In 2011 Cambridge University Press published his e-mail correspondence with scientists around the world in a 600-page tome entitled *Coming of Age with Quantum Information*. As it chronicles the birth pangs of QBism, it offers a glimpse of how theoretical physics is created by real-life, warm-blooded human beings, not the two-dimensional creatures of Wikipedia. The book also documents Fuchs's conviction, contrary to most scientists, that philosophy matters, not only in the way in which it influences physics but also in the manner in which it is informed by the profound insights of physics—or should be.

**Possible Probables**

Fuch's openness to philosophical concerns becomes clear when you consider how QBism forces us to reconsider what is meant by probability. Probability is like "time": we know what it is, until we are asked to define it. Sure, the 50 percent probability of throwing heads with a fair coin implies something about 100 tosses, but how does that intuition help to make sense of the proposition that "the probability of rain this evening is 60 percent" or President Barack Obama's 55/45 assessment, before the event, of the probability of success for the bin Laden operation?

Over the past three centuries two competing definitions of probability have been developed, each with countless variants. The modern, orthodox alternative, called frequentist probability, defines an event's probability as its relative frequency in a series of trials. This number is claimed to be objective and verifiable, as well as directly applicable to scientific experiments. The typical example is the coin toss: in a large number of throws, about half will be heads, so the probability for finding heads is approximately ½. (To avoid the vague words "large," "about" and "approximately," the definition is refined to require an infinite number of tosses, in which case the probability takes on its exact value of ½. Unfortunately, the value also becomes unverifiable at this point and thereby

loses its claim to objectivity.) Applying this definition to weather prediction, one might count real or simulated weather patterns, but as far as President Obama's hunch is concerned, the frequency interpretation is useless—the bin Laden mission was manifestly irreproducible.

The older point of view, Bayesian probability, is named after 18th-century English clergyman Thomas Bayes, whose ideas were perfected and promulgated by French physicist Pierre-Simon Laplace. In contrast to frequentist probability, Bayesian probability is subjective, a measurement of the *degree of belief* that an event will occur. It is a numerical measure of how an agent would bet on the outcome of the event. In simple cases such as coin tosses, frequentist and Bayesian probabilities agree. For the prediction of the weather or of the outcome of a military action, the Bayesian, unlike the frequentist, is at liberty to combine quantitative statistical information with intuitive estimates based on previous experience.

The Bayesian interpretation easily deals with single cases, about which frequentism is silent, and avoids the pitfalls of infinity, but its real power is more specific. On the basis of this interpretation, probability assignments are subject to change because degrees of belief are not fixed. A weather forecaster who is a frequentist would have no trouble calculating the likelihood of rain if the region has had a stable, predictable climate for many years. But in the case of a sudden change, such as a drought, for which there are little data, a Bayesian forecaster is better equipped to account for the new information and the climate condition.

Central to the theory is an explicit formula, called Bayes's rule, for calculating the effect of new information on the estimate of a probability. For example, when a patient is suspected of having cancer, the physician assigns an initial probability, called the prior, based on data such as the known incidence of the disease in the general population, the patient's family history and other relevant factors. On receiving the patient's test results, the doctor then updates this probability using Bayes's rule. The resulting number is no more and no less than the doctor's personal degree of belief.

Most physicists profess their faith in frequentist rather than Bayesian probability, simply because they have been taught to shun subjectivity. But

when it comes to making a prediction, the Bayesian approach rules, says Marcus Appleby, a mathematician at the University of Sydney, who credits Fuchs with convincing him of the significance of Bayesian probability.

Appleby points out that we would consider it crazy to bet in a lottery after learning that the same person has won it every week for 10 years even though a strict frequentist would argue that the results of prior draws have no effect on future outcomes. In practice, no one would ignore the outcome of the previous weeks. Instead the commonsense move would be to adopt the Bayesian viewpoint, update our knowledge and act according to the best available evidence.

**Rewriting Quantum Rules**

Although Qbism negates the reality of the wave function, it is not some nihilistic theory that negates all reality, emphasizes QBism co-author Schack. The quantum system examined by an observer is indeed very real, he notes. Philosophically, Mermin says, QBism suggests a split or boundary between the world in which the observer lives and that person’s experience of that world, the latter described by a wave function.

Mathematically, Fuchs recently made an important discovery that could help cement QBism’s stake as a valid interpretation of probability and quantum theory. The finding has to do with an empirical formula, known as the Born rule, that allows experimentalists to use the wave function of a system to calculate the probability of observing a quantum event in that system. (In technical terms, the Born rule says that we can measure the likelihood of finding a quantum system having property $X$ by taking the square of the magnitude of the wave function assigned to $X$.) Fuchs demonstrated that the Born rule could be rewritten almost entirely in terms of the language of probability theory, without referring to a wave function. The Born rule used to be the bridge that connected wave functions to the results of experiments; now Fuchs has shown that we can predict the results of experiments using probabilities alone.

For Fuchs, the new expression of the Born rule provides another hint that the wave function is just a tool that tells observers how to calculate their personal beliefs, or probabilities, about the quantum world around them. “The Born rule in these lights is an addition to Bayesian probability, not in

the sense of a supplier of some kind of more-objective probabilities, but in the sense of giving extra rules to guide the agent's behavior when he interacts with the physical world," Fuchs writes.

The simplicity of the new equation is striking. Except for one tiny detail, it resembles the law of total probability, the logical requirement that the probabilities for all possible outcomes add up to unity—for example, for a coin flip, the probability of landing on heads (½) plus the probability of landing on tails (½) must equal 1. The deviant detail—the one and only reference to quantum mechanics in this prescription for how to calculate probabilities in quantum theory—is the appearance of *d*, the quantum dimension of the system. Dimension in this sense does not refer to length or width but to the number of states a quantum system can occupy. For instance, a single electron that can have either spin up or spin down would have a quantum dimension of 2.

Fuchs points out that quantum dimension is an intrinsic, irreducible attribute that characterizes the "quantum nature" of a system, in the same way that the mass of an object characterizes its gravitational and inertial properties. Although d is implicit in all quantum-mechanical calculations, its explicit, prominent appearance in a fundamental equation is unprecedented. With the Born rule in its new coat, Fuchs hopes to have discovered the key to a new perspective on quantum mechanics. "I toy," he confesses, "with the idea of [the Born rule] being the most significant 'axiom' of all for quantum theory."

### A New Reality

One of the criticisms of QBism is that it is unable to explain complex macroscopic phenomena in terms of more primitive microscopic ones in the way that conventional quantum mechanics does. The most direct way of meeting that challenge is for QBism to succeed in its stated aim of building the standard theory of quantum mechanics on a foundation of new, compelling assumptions.

That goal has yet to be reached, but even now QBism offers a new view of physical reality. By interpreting the wave function as personal degrees of belief, it gives precise, mathematical meaning to Bohr's intuition that "physics concerns what we can say about nature." And proponents of

QBism embrace the notion that until an experiment is performed, its outcome simply does not exist.

Before an electron is actually observed, for example, the particle exists and moves, but it does not have a speed or a position. Those “properties” have meaning only *following* an observation; it is the act of measurement that creates them. As participants in quantum experiments, we thus become active contributors to the ongoing creation of the universe.

--Originally published: Scientific American 24(4); 92-97 (Winter 2015).

# SECTION 4

# Life at the Limits

# Living in a Quantum World

by Vlatko Vedral

According to standard physics textbooks, quantum mechanics is the theory of the microscopic world. It describes particles, atoms and molecules but gives way to ordinary classical physics on the macroscopic scales of pears, people and planets. Somewhere between molecules and pears lies a boundary where the strangeness of quantum behavior ends and the familiarity of classical physics begins. The impression that quantum mechanics is limited to the microworld permeates the public understanding of science. For instance, Columbia University physicist Brian Greene writes on the first page of his hugely successful (and otherwise excellent) book *The Elegant Universe* that quantum mechanics “provides a theoretical framework for understanding the universe on the smallest of scales.” Classical physics, which comprises any theory that is not quantum, including Albert Einstein’s theories of relativity, handles the largest of scales.

Yet this convenient partitioning of the world is a myth. Few modern physicists think that classical physics has equal status with quantum mechanics; it is but a useful approximation of a world that is quantum at all scales. Although quantum effects may be harder to see in the macroworld, that is a consequence not of size per se but of the way that quantum systems interact with one another.

For many years experimentalists could not confirm that quantum behavior persists on a macroscopic scale. Today they routinely do. These effects are more pervasive than anyone ever suspected. They may even operate in the cells of our bodies.

Quantum behavior eludes visualization and common sense. It forces us to rethink how we look at the universe and accept a new and unfamiliar

picture of our world.

### A Tangled Tale

To a quantum physicist, classical physics is a black-and-white image of a Technicolor world. In the old textbook view, the rich hues become washed out as size increases. Individual particles are quantum; bulk matter is classical. But clues that size is not the determining factor go back to the famous thought experiment of Schrödinger's cat.

Erwin Schrödinger came up with his morbid scenario in 1935 to illustrate how the microworld and macroworld couple to each other, preventing arbitrary lines from being drawn between them. Quantum mechanics says that a radioactive atom can be both decayed and not decayed at the same time. If the atom is linked to a bottle of cat poison, so that the cat dies if the atom decays, then the animal exists in the same quantum limbo as the atom. The weirdness of the one infects the other. Size does not matter. The puzzle was why cat owners only ever see their pets as alive or dead.

In the modern point of view, the world looks classical because the complex interactions that an object has with its surroundings conspire to conceal quantum effects from our view. Information about a cat's state of health, for example, rapidly leaks into its environment in the form of photons and an exchange of heat. The leakage of information causes the combinations of different classical states (such as both dead and alive) that are so distinctive of quantum phenomena to dissipate, through a process known as decoherence.

Larger things tend to be more susceptible to decoherence than smaller ones. That is how physicists usually justify regarding quantum mechanics as a theory of the microworld. But in many cases, the information leakage can be slowed or stopped—and the quantum world then reveals itself to us in all its glory.

The quintessential quantum effect is entanglement, which binds together individual particles into an indivisible whole. A classical system is always divisible, at least in principle; whatever collective properties it has arise from the combined properties of its components. But an entangled system cannot be dissected this way. One strange consequence of entanglement is

that entangled particles behave as a single entity, even when they are far apart. Einstein famously called the effect “spooky action at a distance.”

Experiments commonly entangle pairs of elementary particles, such as electrons. You can imagine such particles as small spinning tops that rotate, clockwise or counterclockwise, around axes that can point in any direction: horizontally, vertically or at some angle in between. To measure a particle’s spin, you must first choose a direction. Then you can test whether the particle spins in that direction.

Suppose electrons actually behaved classically. You might set up one electron to spin in the horizontal clockwise direction and the other in the horizontal counterclockwise direction; that way, the sum of their spins is zero. Their axes remain fixed in space, and when you make a measurement, the outcome depends on whether the direction you choose aligns with the particles’ horizontal axis. If it does, then you see them spinning in opposite directions, but if you measure them vertically, neither appears to have any spin at all.

In real experiments, however, the outcome is astonishingly different because real electrons follow quantum rules. You can set up the particles to have a total spin of zero even when you have not specified what their individual spins are. Then, when you measure one of the particles, you will see it spinning clockwise or counterclockwise at random. It is as though the particle decides which way to spin for itself. Nevertheless, no matter which direction you choose to measure the electrons (providing it is the same for both), they will always spin in opposite ways, one clockwise and the other counterclockwise.

How do they know to do so? That remains utterly mysterious. What is more, if you measure one particle horizontally and the other vertically, you will still detect non-zero spin for each, which implies that the particles have no fixed axes of rotation. Classical physics cannot explain this.

---

## Observing the Observer

The idea that quantum mechanics applies to everything in the universe, even to us humans, can lead to some strange conclusions. Consider a variant of the iconic Schrödinger cat thought experiment (described in the main text above). Nobel laureate Eugene P. Wigner came up with this version in 1961, and David Deutsch of the University of Oxford elaborated on it in 1986.

Suppose that a very able experimental physicist, Alice, puts her friend Bob inside a room with a cat, a radioactive atom and cat poison that gets released if the atom decays. The point of having a human there is that we can communicate with him. (Getting answers from cats is not that easy.) As far as Alice is concerned, the atom enters into a state of being both decayed and not decayed, so that the cat is both dead and alive. Bob, however, can directly observe the cat and sees it as one or the other. Alice slips a piece of paper under the door asking Bob whether the cat is in a definite state. He answers, “Yes.”

Note that Alice does *not* ask whether the cat is dead or alive because for her that would force the outcome or, as physicists say, “collapse” the state. She is content observing that her friend sees the cat as either alive or dead and does not ask which it is.

Because Alice avoided collapsing the state, quantum theory holds that slipping the paper under the door was a reversible act. She can undo all the steps she took. If the cat was dead, it would now be alive, the poison would be in the bottle, the particle would not have decayed and Bob would have no memory of ever seeing a dead cat.

Yet one trace remains: Bob’s answer on the paper. Alice can undo the observation in a way that does not force his writing to vanish. Written proof remains that Bob observed the cat as definitely alive or dead.

That leads to a startling conclusion. Alice was able to reverse the observation because, as far as she was concerned, she avoided collapsing the state; to her, Bob was in just as indeterminate a state as the cat. But the friend inside the room thought the state did collapse. Bob did see a definite outcome; his answer is proof of it. In this way, the experiment demonstrates two seemingly contradictory principles. Alice thinks that quantum mechanics applies to macroscopic objects: not just cats but also Bobs can be in quantum limbo. Bob thinks that cats are only either dead or alive.

Doing such an experiment with an entire human being would be daunting, but physicists can accomplish much the same with simpler systems. Anton Zeilinger and his colleagues at the University of Vienna take a photon and bounce it off a large mirror. If the photon is reflected, the mirror recoils, but if the photon is transmitted, the mirror stays still. The photon plays the role of the decaying atom; it can exist simultaneously in more than one state. The mirror, made up of billions of atoms, acts as the cat and as Bob. Whether it recoils or not is analogous to whether the cat lives or dies and is seen to live or die by Bob. The process can be reversed by reflecting the photon back at the mirror.

On smaller scales, teams led by Rainer Blatt of the University of Innsbruck and by David J. Wineland of the National Institute of Standards and Technology in Boulder, Colo., have reversed the measurement of vibrating ions in an ion trap. Wineland was awarded the 2012 Nobel Prize in Physics for such advances in quantum experimentation.

Like those of Erwin Schrödinger and Albert Einstein before them, Wigner and Deutsch’s thought experiment suggests that physicists have yet to grasp quantum mechanics in any deep way. For decades most physicists scarcely cared because the foundational issues had no effect on practical applications of the theory. But now that we can perform these experiments for real, the task of under standing quantum mechanics has become more urgent.

— V.V.

---

## Quantum Salt

Physicists used to think that distinctive quantum phenomena would operate only at the level of individual particles; great big clusters of particles would behave classically. Recent experiments show otherwise. For example, the atoms in a salt crystal typically point every which way (*below left*) and line up when physicists apply a magnetic field. But they line up faster than they would if only classical physics operated (*below center*). Evidently the

quantum phenomenon of entanglement—the "spooky action" that coordinates the properties of far-flung particles—is helping bring them into line (*below right*). The role of entanglement is revealed by a measure of the crystal's magnetic properties (graph).

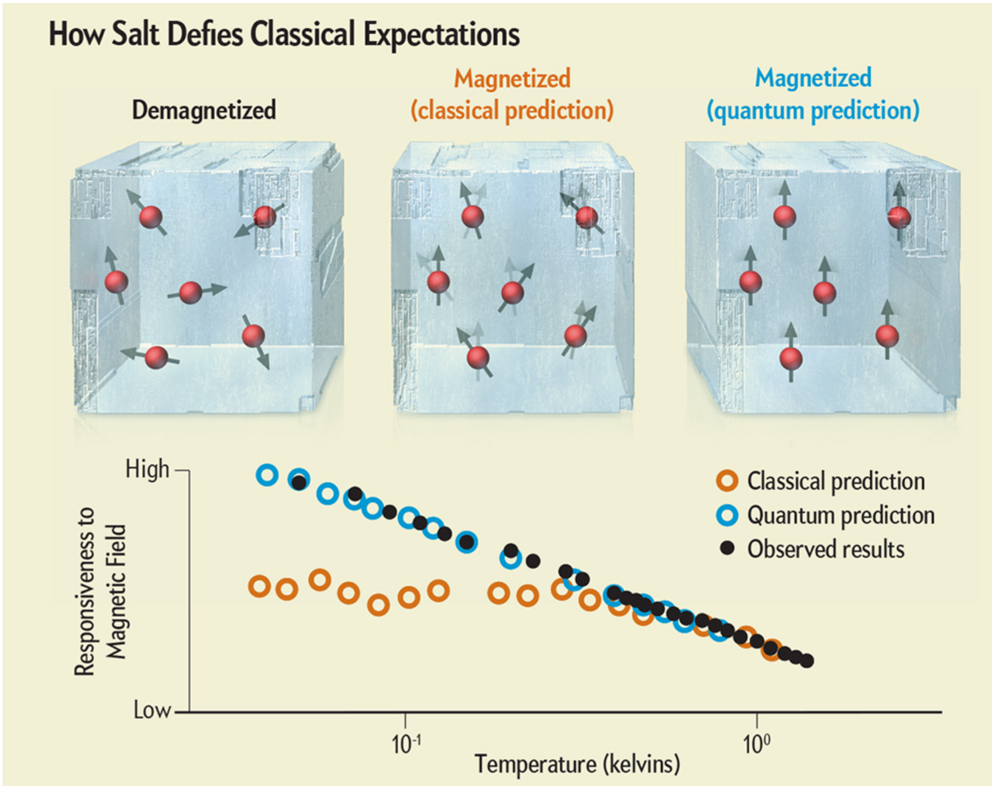



*Illustration by George Retseck, Graphic by Jen Christiansen*
*Source for Graph: "Entangled Quantum State of Magnetic Dipoles," by S. Ghosh et al., In Nature, Vol. 425; September 4, 2003.*

## Acting as One

Most demonstrations of entanglement involve at most a handful of particles. Larger batches of entangled particles are harder to isolate; they tend to form unwanted entanglements with stray particles, obscuring their original interconnections. Too much information leaks out to the environment, and the system then behaves classically. The difficulty of preserving entanglement is a major challenge for those of us seeking to exploit these novel effects for practical use, such as quantum computers.

But a neat experiment in 2003 proved that larger systems, too, can remain entangled when the leakage is reduced or somehow counteracted. Gabriel Aeppli, then at University College London, and his colleagues put a piece of lithium fluoride salt in an external magnetic field. You can think of the atoms in the salt as little spinning magnets that try to align themselves with the external field. The atoms exert forces on one another, which act as a

kind of peer pressure to bring them into line more quickly. As the researchers varied the strength of the magnetic field, they measured how quickly the atoms became aligned.

They found that the atoms responded much faster than the strength of their mutual interactions would suggest. Evidently some additional effect was helping the atoms act in unison. The researchers argued that entanglement was the culprit. If so, the $10^{20}$ atoms of the salt formed a hugely entangled state.

To avoid the confounding effects of the random motions associated with heat energy, Aeppli's team performed the experiments at extremely low temperatures—a few millikelvins. Since then, however, Alexandre Martins de Souza of the Brazilian Center for Physics Research in Rio de Janeiro and his colleagues have discovered macroscopic entanglement in materials such as copper carboxylate at room temperature and higher. In these systems, the interaction among particle spins is strong enough to resist thermal chaos.

In other cases, an external force wards off thermal effects. Physicists have seen entanglement in systems of increasing temperature and size, from ions trapped by electromagnetic fields to ultracold atoms in lattices to superconducting quantum bits.

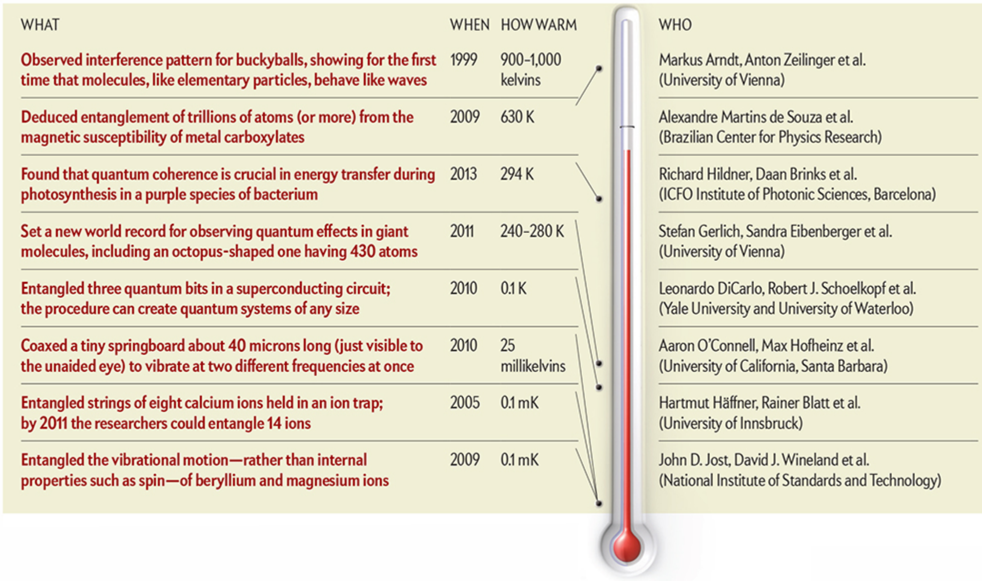
These systems are analogous to Schrödinger's cat. Consider an atom or ion. Its electrons can exist close to the nucleus or farther away—or both at the same time. Such an electron acts like the radioactive atom that has either decayed or not decayed in Schrödinger's thought experiment. Independently of what the electron is doing, the entire atom can be moving, say, left or right. This motion plays the role of the dead or alive cat. By manipulating the atom with laser light, physicists can couple the two properties. Then, if the electron is close to the nucleus, the atom moves to the left; if the electron is farther away, the atom moves to the right. The state of the electron is entangled with the movement of the atom, just as radioactive decay is entangled with the state of Schrödinger's cat.

By scaling up this basic idea, experimentalists have entangled huge numbers of atoms into states that classical physics would deem impossible. And if solids can be entangled even when they are large and warm, we can

ask whether the same might be true of a very special kind of large, warm system: life.

Entanglement Heats Up

Quantum effects are not limited to cold subatomic particles. They also show up in experiments on larger and warmer systems.



| WHAT | WHEN | HOW WARM | WHO |
|---|---|---|---|
| Observed interference pattern for buckyballs, showing for the first time that molecules, like elementary particles, behave like waves | 1999 | 900–1,000 kelvins | Markus Arndt, Anton Zeilinger et al. (University of Vienna) |
| Deduced entanglement of trillions of atoms (or more) from the magnetic susceptibility of metal carboxylates | 2009 | 630 K | Alexandre Martins de Souza et al. (Brazilian Center for Physics Research) |
| Found that quantum coherence is crucial in energy transfer during photosynthesis in a purple species of bacterium | 2013 | 294 K | Richard Hildner, Daan Brinks et al. (ICFO Institute of Photonic Sciences, Barcelona) |
| Set a new world record for observing quantum effects in giant molecules, including an octopus-shaped one having 430 atoms | 2011 | 240–280 K | Stefan Gerlich, Sandra Eibenberger et al. (University of Vienna) |
| Entangled three quantum bits in a superconducting circuit; the procedure can create quantum systems of any size | 2010 | 0.1 K | Leonardo DiCarlo, Robert J. Schoelkopf et al. (Yale University and University of Waterloo) |
| Coaxed a tiny springboard about 40 microns long (just visible to the unaided eye) to vibrate at two different frequencies at once | 2010 | 25 millikelvins | Aaron O'Connell, Max Hofheinz et al. (University of California, Santa Barbara) |
| Entangled strings of eight calcium ions held in an ion trap; by 2011 the researchers could entangle 14 ions | 2005 | 0.1 mK | Hartmut Häffner, Rainer Blatt et al. (University of Innsbruck) |
| Entangled the vibrational motion—rather than internal properties such as spin—of beryllium and magnesium ions | 2009 | 0.1 mK | John D. Jost, David J. Wineland et al. (National Institute of Standards and Technology) |

*Illustration by George Retseck*

## Schrödinger's Birds

European robins are crafty little birds. Every year they migrate from Scandinavia to the warm plains of equatorial Africa and return in the spring, when the weather up north becomes more tolerable. The robins navigate this round-trip of some 13,000 kilometers with natural ease.

People have long wondered whether birds and other animals might have some built-in compass. In the 1970s the husband-wife team of Wolfgang and Roswitha Wiltschko of the University of Frankfurt in Germany caught robins that had been migrating to Africa and put them in artificial magnetic fields. Oddly, the robins seemed oblivious to a reversal of the magnetic field direction; they could not tell north from south. The birds did, however, respond to the inclination of the earth's magnetic field—that is, the angle that the field lines make with the surface. That is all they need to navigate. Interestingly, blindfolded robins did not respond to a magnetic field at all, indicating that they somehow sense the field with their eyes.

In 2000 Thorsten Ritz, a physicist who has a passion for migratory birds and is now at the University of California, Irvine, and his colleagues proposed that entanglement is the key. In their scenario, which builds on the previous work by Klaus J. Schulten of the University of Illinois at Urbana-Champaign, a bird's eye has a type of molecule in which two electrons form an entangled pair with zero total spin. Such a situation simply cannot be mimicked with classical physics. When this molecule absorbs visible light, the electrons get enough energy to separate and become susceptible to external influences, including the earth's magnetic field. If the magnetic field is inclined, it affects the two electrons differently, creating an imbalance that changes the chemical reaction that the molecule undergoes. Chemical pathways in the eye translate this difference into neurological impulses, ultimately creating an image of the magnetic field in the bird's brain.

Although the evidence for Ritz's mechanism is circumstantial, Christopher T. Rodgers of the University of Oxford and Kiminori Maeda, now at Saitama University in Japan, studied molecules similar to Ritz's in the laboratory and showed that electron entanglement does make these molecules sensitive to magnetic fields.

Calculations that my colleagues and I have done suggest that quantum effects persist in a bird's eye for around 100 microseconds—which, in this context, is a long time. The record for an artificially engineered electron-spin system is about 50 microseconds. We do not yet know how a natural system could preserve quantum effects for so long, but the answer could give us ideas for how to protect quantum computers from decoherence.

Photosynthesis, whereby plants convert sunlight into chemical energy, may also involve entanglement. Incident light ejects electrons inside plant cells, and these electrons all need to find their way to the same place: the chemical reaction center where they can deposit their energy and set off the reactions that fuel plant cells. Classical physics fails to explain the near-perfect efficiency with which they do so.

Experiments by several groups, such as Graham R. Fleming and his colleagues at the University of California, Berkeley, Mohan Sarovar of Sandia National Laboratories and Gregory D. Scholes of the University of Toronto, suggest that quantum mechanics accounts for the high efficiency

of the process. In a quantum world, a particle does not just have to take one path at a time; it can take all of them simultaneously. The electromagnetic fields within plant cells can cause some of these paths to cancel one another and others to reinforce mutually.

The net effect is to reduce the chance that the electron will take a wasteful detour and to increase the odds that it will head straight to the reaction center. The entanglement would last only a fraction of a second and would involve molecules that have no more than about 100,000 atoms. Do any instances of larger and more persistent entanglement exist in nature? We do not know, but the question is exciting enough to stimulate an emerging discipline: quantum biology.

**The Meaning of It All**

Schrödinger thought the prospect of cats that were both alive and dead was absurd; any theory making such a prediction must surely be flawed. Generations of physicists shared this discomfort and thought that quantum mechanics would cease to apply at a still larger scale. In the 1980s Roger Penrose of Oxford suggested that gravity might cause quantum mechanics to give way to classical physics for objects more massive than 20 micrograms, and a trio of Italian physicists— GianCarlo Ghirardi and Tomaso Weber of the University of Trieste and Alberto Rimini of the University of Pavia—proposed that large numbers of particles spontaneously behave classically.

But experiments now leave little room for such processes to operate. The division between the quantum and classical worlds appears not to be fundamental. It is just a question of experimental ingenuity, and few physicists now think that classical physics will ever really make a comeback at any scale. If anything, the general belief is that if a deeper theory ever supersedes quantum physics, it will show the world to be even more counterintuitive than anything we have seen so far.

The fact that quantum mechanics applies on all scales forces us to confront the theory's deepest mysteries. We cannot simply write them off as mere details that matter only in the realm of the minuscule. For instance, space and time are two of the most fundamental classical concepts, but according to quantum mechanics they are secondary. The entanglements are

primary. They interconnect quantum systems without reference to space and time. If there were a dividing line between the quantum and the classical worlds, we could use the space and time of the classical world to provide a framework for describing quantum processes. But without such a dividing line—and, indeed, without a truly classical world—we lose this framework. We must explain space and time as somehow emerging from fundamentally spaceless and timeless physics.

That insight, in turn, may help us reconcile quantum physics with that other great pillar of physics, Einstein's general theory of relativity, which describes the force of gravity in terms of the geometry of spacetime. General relativity assumes that objects have well-defined positions and never reside in more than one place at the same time—in direct contradiction with quantum physics. Many physicists, such as Stephen Hawking of the University of Cambridge, have argued that relativity theory must give way to a deeper theory in which classical spacetime emerges out of quantum entanglements through the process of decoherence.

Gravity might not even be a force in its own right but rather the residual noise emerging from the quantum fuzziness of the other forces in the universe. This idea of "induced gravity" goes back to nuclear physicist Andrei Sakharov in the 1960s. If true, it would not only demote gravity from the status of a fundamental force but also suggest that efforts to "quantize" gravity are misguided. Gravity may not even exist at the quantum level.

The implications of macroscopic objects such as our human bodies being in quantum limbo is mind-blowing enough that we physicists are still in an entangled state of confusion and wonderment.

--Originally published: Scientific American 24(4) 98-103 (Winter 2015).

# Limits of Intelligence

by Douglas Fox

It is humbling to think that a honeybee, with its milligram-size brain, can perform tasks such as navigating mazes and landscapes on a par with mammals. A honeybee may be limited by having comparatively few neurons, but it surely seems to squeeze everything it can out of them. At the other extreme, an elephant, with its five-million-fold larger brain, suffers the inefficiencies of a sprawling Mesopotamian empire. Signals take more than 100 times as long to travel between opposite sides of its brain—and also from its brain to its foot, forcing the beast to rely less on reflexes, to move more slowly, and to devote 97 percent of the neurons in its brain to the cerebellum, which coordinates each step.

We humans may not occupy the dimensional extremes of elephants or honeybees, but what few people realize is that the laws of physics place tough constraints on our mental faculties as well. Anthropologists have speculated about anatomic roadblocks to brain expansion—for instance, whether a larger brain could fit through the birth canal of a bipedal human. If we assume, though, that evolution (or surgeons) can solve the birth-canal problem, then we are led to the cusp of some even more profound questions.

One might think, for example, that evolutionary processes could increase the number of neurons in our brain or boost the rate at which those neurons exchange information. Such changes could, in principle, make us smarter. But several recent lines of investigation, if taken together and followed to their logical conclusion, seem to suggest that such avenues for improvement would soon be blocked by physical limits. Ultimately those limits trace back to the very nature of neurons and the statistically noisy ways in which they communicate by chemical exchanges. “Information, noise and energy are inextricably linked,” says Simon Laughlin, a theoretical neuroscientist

at the University of Cambridge. “That connection exists at the thermodynamic level.”

Do the laws of thermodynamics, then, impose a limit on neuron-based intelligence, one that applies universally, whether in birds, primates, porpoises or praying mantises? This question apparently has never been asked in such broad terms, but the scientists interviewed for this article generally agree that it is a question worth contemplating.

“It’s a very interesting point,” says Vijay Balasubramanian, a physicist who studies neural coding of information at the University of Pennsylvania. “I’ve never even seen this point discussed in science fiction.”

Intelligence is of course a loaded word: it is hard to measure and even to define. Still, it seems fair to say that by most metrics, humans are the most intelligent animals on earth. But as our brain has evolved, has it approached a hard limit to its ability to process information? Could there be some physical limit to the evolution of neuron-based intelligence—and not just for humans but for all of life as we know it?

**That Hungry Tapeworm in Your Head**

The most intuitively obvious way in which brains could get more powerful is by growing larger. And indeed, the possible connection between brain size and intelligence has fascinated scientists for more than 100 years. Biologists spent much of the late 19th century and the early 20th century exploring universal themes of life, mathematical laws related to body mass—and to brain mass in particular—that run across the animal kingdom. One advantage of size is that a larger brain can contain more neurons, which should enable it to grow in complexity as well. But it was clear even then that brain size alone did not determine intelligence: a cow’s brain is well over 100 times the size of a mouse’s, but the cow isn’t any smarter.

Instead brains seem to expand with body size in order to carry out more trivial functions. Bigger bodies might, for example, impose more neural housekeeping chores unrelated to intelligence, such as monitoring larger numbers of tactile nerves, processing signals from bigger retinas and controlling more muscle fibers.

Eugene Dubois, the Dutch anatomist who discovered the skull of *Homo erectus* in Java in 1892 , wanted a way to estimate the intelligence of animals based on the size of their fossil skulls, so he worked to define a precise mathematical relation between the brain size and body size of animals—under the assumption that animals with disproportionately large brains would also be smarter. Dubois and others amassed an ever growing database of brain and body weights. One classic treatise reported the body, organ and gland weights of 3,690 animals, from wood roaches and yellow-billed egrets to slugs and three-toed sloths.

Dubois's successors found that mammals' brains expand more slowly than their bodies do and generally reach about the ¾ power of body mass. So a muskrat, which has a body 38 times as large as a mouse's, carries a brain only about eight times as big. From that insight came the tool that Dubois had sought: the encephalization quotient, a ratio of the average brain mass actually measured for a species to the mass predicted by a version of the power law . Humans have a quotient of 7.5, meaning our brains are 7.5 times as large as the law predicts. Bottlenose dolphins sit at 5.3; monkeys hover as high as 4.8; and oxen, unsurprisingly, lumber around at 0.5. In short, intelligence may depend on the amount of neural reserve left over after the brain's menial chores, such as minding skin sensations, are accounted for. Or to boil it down even more: when it comes to intelligence, bigger may be better, at least in superficial ways.

As brains expanded in mammals and birds, the organs almost certainly benefited from economies of scale. There is, for starters, the network effect: the more neural pathways that connect a pair of neurons, the more information each signal implicitly carries. That's one reason that neurons in larger brains can get away with firing fewer times a second.

Meanwhile, however, a competing trend may have kicked in. Balasubramanian argues that "it is very likely that there is a law of diminishing returns" to increasing intelligence indefinitely by adding new brain cells. Size carries burdens with it, the most obvious one being added energy consumption. In humans, the brain is already the hungriest part of our body: at 2 percent of our body weight, this greedy little tapeworm of an organ wolfs down 20 percent of the calories that we expend at rest. In newborns, it's an astounding 65 percent.

## Staying in Touch

Much of the energetic burden of brain size comes from the organ's communication networks. In the human cortex, communications eats up four fifths of the energy expended. But it appears that as size increases, neuronal connectivity also becomes more challenging for subtler, structural reasons. In fact, even as biologists kept collecting data on brain mass in the early to mid-20th century, they also took on the more daunting challenge of defining the "design principles" of brains and explaining how the principles work in brains of vastly different sizes.

A typical neuron extends an elongated tail called the axon. At its end, the axon branches out, and the tips of the branches form synapses, or contact points, with other cells. Like telegraph wires, axons can connect distant parts of the brain, but they can also bundle together to form nerves that link the central nervous system to far-flung parts of the body.

Pioneering neuroscientists measured the diameter of axons under microscopes. They counted the size and density of nerve cells, as well as the number of synapses per cell. They surveyed hundreds—sometimes thousands—of cells per brain in dozens of species. Eager to refine their mathematical curves by extending them to ever larger beasts, they even found ways to extract intact brains from whale carcasses. The five-hour process involved the use of a two-man lumberjack saw, an ax, a chisel and brute strength to open the top of the skull like a giant can of beans.

These studies revealed that as brains expand in size from species to species, several subtle but probably unsustainable changes happen. First, the average size of nerve cells increases. This phenomenon allows the neurons to connect to more and more of their peers as the overall number of neurons in the brain increases. But because larger cells pack into the cerebral cortex less densely, the distance between cells increases, as does the length of axons required to connect them. What's more, longer axons mean longer times for signals to travel between cells. To compensate for the distance, axons have to get thicker—they can then carry signals faster.

Researchers have also found that as brains get bigger from species to species, they are divided into a larger and larger number of distinct areas. You can see those areas if you stain brain tissue and view it under a

microscope: patches of the cortex turn different colors. These areas often correspond with specialized functions, such as speech comprehension or face recognition. The specialization unfolds in another dimension as well: equivalent areas in the left and right hemispheres take on separate functions—for example, spatial versus verbal reasoning.

For decades this dividing of the brain into more work cubicles was viewed as a hallmark of intelligence. But it may also reflect a more mundane truth, says Mark Changizi, a theoretical neurobiologist at 2AI Labs in Boise, Idaho. Specialization compensates for the connectivity problem that arises as brains get bigger. If a cow's brain had the same design as a mouse's brain, despite having 100 times as many neurons, there is no way the neurons could be as well connected as they are in the mouse. Cows—and other large mammals—solve this problem by segregating like-functioned brain neurons into highly interconnected modules and reducing the numbers of direct long-distance connections among modules.

The specialization of the hemispheres similarly reduces the amount of information that must leap across long, interhemispheric axons from one side of the brain to the other. "All of these seemingly complex things about bigger brains are just the backbends that the brain has to do to satisfy the connectivity problem" as it gets larger, Changizi argues. "It doesn't tell us that the brain is smarter."

Jan Karbowski, a computational neuroscientist at the University of Warsaw in Poland, agrees. "Somehow brains have to optimize several parameters simultaneously, and there must be trade-offs," he says. "If you want to improve one thing, you screw up something else."

What happens, for example, if you expand the corpus callosum (the bundle of axons connecting right and left hemispheres) quickly enough to maintain constant connectivity as brains expand? And what if you further thicken those axons, so the transit delay for signals traveling between hemispheres does not increase as brains expand? The results would not be pretty. The corpus callosum would expand—and push the hemispheres apart—so quickly that any performance improvements would be neutralized.

Recent experiments refining the relation between axon width and conduction speed have brought these trade-offs into stark relief. Neurons do grow larger as brain size increases, Karbowski says, but not quickly enough to prevent a decrease in connectivity. And while axons do thicken as brains expand, it’s not enough to offset the longer conduction delays.

There is a good reason that axons don’t thicken more. Restraining their girth saves the brain both space and energy, Balasubramanian says. Double the width of an axon, and its energy expenditure doubles, but the velocity of its pulses goes up just 40 percent or so.

Even with all of this corner cutting, the volume of white matter (the axons) still grows more quickly than the volume of gray matter (the main body of neurons containing the cell nucleus) as brains increase in size. To put it another way, as brains get bigger, more of their volume is devoted to wiring rather than to the parts of individual cells that do the actual computing. This alone suggests that scaling size up is ultimately unsustainable.

**The Primacy of Primates**

The seemingly intractable inefficiencies that big-brained beings face explain why a cow fails to squeeze any more smarts out of its grapefruit-size brain than a mouse does from its blueberry-size brain. But how is it that humans are so smart? Part of the answer seems to be that evolution has found impressive workarounds at the level of the brain’s building blocks. When Suzana Herculano-Houzel, a neuroscientist at the Federal University of Rio de Janeiro in Brazil, surveyed the number and size of brain cells across 41 mammal species in 2014, she and her colleagues stumbled on to a game changer that probably gives humans an edge.

Herculano-Houzel found that cortical neurons in primates differ in an important way from those in most other mammals. In primates, only a few of these neurons grow much larger as the brain increases in size. These rare oversized neurons may shoulder the burden of keeping things well connected and allow the majority to remain small. This feature allows the brains of large primates (humans included) to stay dense. An owl monkey, for example, has about twice the brain mass of a marmoset, as well as

roughly twice as many neurons. In contrast, a similar doubling of mass in the rodent brain boosts the neuron count by just 60 percent.

"It's a huge difference—one of the things that makes a primate a primate," Herculano-Houzel says. Humans pack 86 billion neurons into 1.4 kilograms of brain, but a rodent that had followed its usual neuron-size scaling law to reach that number of neurons would now have to drag around a brain weighing 45 kilograms. And metabolically speaking, all that brain matter would eat the varmint out of house and home.

Having smaller, more densely packed neurons does seem to have a real impact on intelligence. In 2005 neurobiologists Gerhard Roth and Ursula Dicke, both at the University of Bremen in Germany, reviewed several traits that predict intelligence across species (as measured, roughly, by behavioral complexity) even more effectively than the encephalization quotient does. "The only tight correlation with intelligence," Roth says, "is in the number of neurons in the cortex, plus the speed of neuronal activity," which decreases with the distance between neurons and increases with the degree of myelination of axons. (Myelin is fatty insulation that lets axons transmit signals more quickly.)

If Roth is right, then primates' small neurons confer a double advantage. First, they allow a greater increase in cortical cell number as brains enlarge. Second, they allow faster communication because the cells pack more closely. Elephants are reasonably smart, but their neurons are six times larger than humans' and up to 40 times larger than other mammals', leading to inefficiencies. "The packing density of neurons is much lower," Roth says, "which means that the distance between neurons is larger and the velocity of nerve impulses is much lower."

A growing number of studies have revealed a similar pattern of variation within humans: people who have the quickest lines of communication among their brain areas also seem to be the brightest. One study, reported in 2014 by Emiliano Santarnecchi of Harvard Medical School, used functional magnetic resonance imaging to measure how directly different brain areas talk to one another—that is, whether they talk via a large or a small number of intermediary areas. Santarnecchi found that shorter paths between brain areas, and higher overall network efficiency, correlated with higher IQ.

Edward Bullmore, an imaging neuroscientist at the University of Cambridge, and his collaborators obtained similar results in 2009 using a different approach. They compared working memory (which is the ability to hold several numbers in one's memory at once) among 29 healthy people. The researchers then used magnetoencephalographic recordings from their subjects' scalps to estimate how quickly communication flowed between brain areas. People whose brains exhibited the most direct communication and the fastest neural chatter had the best working memory.

It is a momentous insight. We know that as brains get larger, they save space and energy by limiting the direct connections among regions. The large human brain contains relatively few of these long- distance connections. But Bullmore and Santarnecchi showed that these rare, nonstop connections have a disproportionate influence on smarts: brains that scrimp on resources by cutting just a few of them do noticeably worse. "You pay a price for intelligence," Bullmore concludes, "and the price is that you can't simply minimize wiring."

**Intelligence Design**

If communication among neurons, and between brain areas, is really a major bottleneck that limits intelligence, then when evolution produces smaller neurons that pack together more densely and communicate faster, that should yield smarter brains. Brains might also become more efficient by evolving axons that can carry signals faster over longer distances without getting thicker. But something prevents animals from shrinking neurons and axons beyond a certain point. You might call it the mother of all limitations: the proteins that neurons use to generate electrical pulses are inherently unreliable.

These proteins, known as ion channels, act like tiny valves to open and close pores in the cell membrane of the neuron. Open channels allow ions of sodium, potassium or calcium to flow into or out of neurons, producing the electrical signals by which these cells communicate. But the channels are so minuscule that mere thermal vibrations can flip them open or closed.

If you were to isolate a single ion channel on the surface of a nerve cell and then adjust the voltage across the channel to open or close it, you would find that this protein-activated switch does not flip on and off reliably like

your kitchen light does. Instead it flutters open and shut unpredictably, like a sticky screen door on a windy day. Changing the voltage only influences the *likelihood* that it will open.

This may sound like a horrible evolutionary design flaw, but it's not a bug—it's a feature. Or rather the unavoidable price of having a sensitive, energy-efficient gate. "If you make the spring on the channel too loose," Laughlin explains, "then the noise keeps on switching it"—the screen door in the wind. Cells could use stiffer proteins as channels to dampen that noise, he says, but that would force neurons to spend more energy to control the ion channel. The trade-off means that ion channels are reliable only when many of them are used in parallel to "vote" on whether or not a neuron should generate an impulse.

Here's the rub: voting becomes problematic as neurons get smaller. "When you reduce the size of neurons, you reduce the number of channels that are available to carry the signal," Laughlin says. "And that increases the noise."

In a pair of papers published in 2005 and 2007, Laughlin and his collaborators calculated whether the need to include enough ion channels limits how small axons can be made. The results were startling. "When axons got to be about 150 to 200 nanometers in diameter, they became impossibly noisy," Laughlin says. At that point, an axon contains so few ion channels that the accidental opening of a single channel can spur the axon to deliver a signal even though the neuron did not intend to fire.

The brain's smallest axons probably already hiccup out about six of these accidental spikes per second. Shrink them just a little bit more, and they would blather out more than 100 a second. "Cortical gray matter neurons are working with axons that are pretty close to the physical limit," Laughlin concludes.

This fundamental compromise among information, energy and noise is not unique to biology. It applies to everything from ham radios to computer chips. Transistors act as gatekeepers of electrical signals, just like ion channels do.

For five decades engineers have shrunk transistors steadily, cramming more and more onto chips to produce ever faster computers. Transistors in

the latest chips have features that are just 14 nanometers across. At that size, it becomes very challenging to “dope” silicon uniformly. (Doping is the addition of small quantities of other elements to adjust a semiconductor’s properties.) When features shrink below 10 nanometers, a transistor will be so small that the random presence or absence of a single doping atom of boron could cause it to behave unpredictably.

Engineers might circumvent these limitations by redesigning chips using new technologies. But life cannot start from scratch: it has to work within the framework that has evolved over 600 million years. And during that time a peculiar thing has happened: one particular grand plan has popped up over and over again.

The brains of the honeybee, the octopus, the crow and intelligent mammals look nothing alike at first glance. But if you look at the circuits that underlie tasks such as vision, smell, navigation and episodic memory of event sequences, “very astonishingly they all have absolutely the same basic arrangement,” Roth says. Such convergence in anatomy and physiology usually suggests that a certain evolutionary solution has reached maturity, leaving little room for improvement.

Perhaps, then, life has arrived at an optimal neural blueprint. That blueprint is wired up through a step-by-step choreography in which cells in the growing embryo interact through signaling molecules and physical nudging. The blueprint may be so entrenched by evolution as to rule out any major changes in plan.

**Bees Do It**

So have humans reached the physical limits of how complex our brain can be, given the building blocks that are available to us? Work by Herculano-Houzel suggests that we have already jumped through a major evolutionary hoop just to obtain the brain we have now.

It all comes down to calories. Animals can spend only so many hours eating per day. Primates thus faced a critical tradeoff as they evolved larger brains, she says. They could expend their limited calories on a larger, more powerful body—or on a smarter brain. Gorillas, orangutans and chimpanzees maxed out their calories with various combinations of big, strong bodies and brains containing 20 to 40 billion neurons. Those brains

consume around 9 percent of the total calories that they burn—which means they must spend up to eight hours a day foraging.

Humans, in contrast, sport brains packed with 86 billion neurons—and we devote a whopping 20 percent of our calories to feeding our heads. We can afford such an extravagant caloric luxury, Herculano-Houzel believes, only because our species developed a unique technology: the cooking fire.

Around 1.5 million years ago our ancestors began using fire to transform food. “That allows a jump in the amount of calories that you can get from your food that no other practice can achieve,” Herculano-Houzel says. Cooking makes it easier to digest plant foods and to extract calorie-dense fat from animal carcasses—for example, by stewing bones to extract marrow. It seems an unlikely coincidence that around the time our human ancestors conquered fire, they also finally broke through the caloric barrier and jumped from brains of perhaps 40 billion brain neurons (*Homo habilis*) to 60 billion neurons (*Homo erectus*), and finally to 86 billion. Were it not for cooking, she says, “we would not be here.”

But what about future human evolution? Laughlin doubts that there is any hard limit on brain function the way there is on the speed of light. “It’s more likely you just have a law of diminishing returns,” he says. “It becomes less and less worthwhile the more you invest in it.”

The human mind, however, may have better ways of expanding without the need for further biological evolution. Through social interaction and language, we humans have learned to pool our intelligence into collective smarts.

And then there is technology. For millennia written language has enabled us to store information outside our body, effectively extending the capacity of our brains. One could argue that the Internet is the ultimate consequence of this trend toward outward expansion of intelligence. In a sense it could be true, as some say, that the Internet makes you stupid: collective human intelligence—culture and computers—may have reduced the impetus for evolving greater individual smarts.

---

Why We Probably Cannot Get Much Smarter

Miniaturization is just one of several evolutionary tweaks that could, in principle, enhance our intelligence and at the same time carry disadvantages and run into

thermodynamic hurdles. Perhaps we are already close to being as smart as a neuron-based intelligence can be.

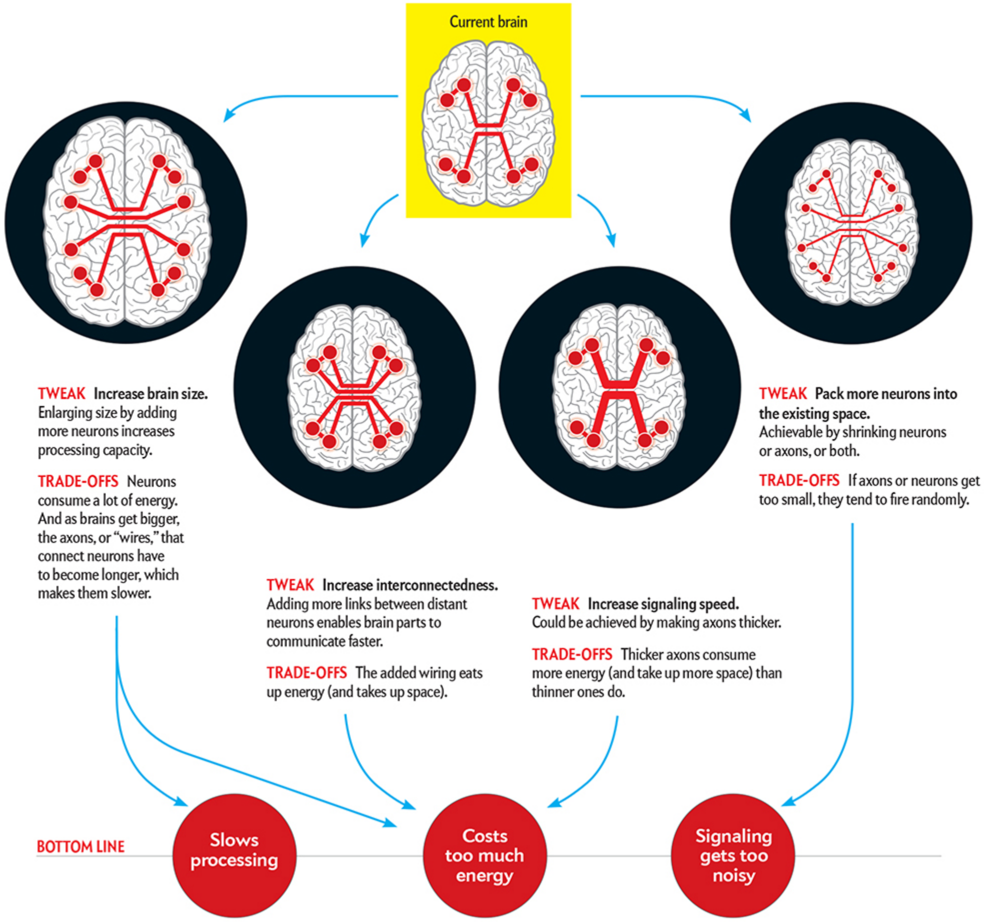



*Illustration by Brown Bird Design*

---

--Originally published: Scientific American 24(4); 104-111 (Winter 2015).